CAMBRIDGE
UNIVERSITY PRESS

## Research Paper

# The bivariate gas–stellar mass distributions and the mass functions of early- and late-type galaxies at $z \sim 0$

Aldo Rodríguez-Puebla[1] ⬥, A. R. Calette[1], Vladimir Avila-Reese[1], Vicente Rodriguez-Gomez[2] and Marc Huertas-Company[3,4,5,6]

[1]Instituto de Astronomía, Universidad Nacional Autónoma de México, A. P. 70-264, 04510, México, D.F., México, [2]Instituto de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, A. P. 72-3, 58089 Morelia, México, [3]Sorbonne Universitè, Observatoire de Paris, Universitè PSL, CNRS, LERMA, F-75014, Paris, France, [4]Univeristé de Paris, 5 Rue Thomas Mann - 75013, Paris, France, [5]Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain and [6]Instituto de Astrofíisica de Canarias, E-38200 La Laguna, Tenerife, Spain

## Abstract

We report the bivariate HI- and $H_2$-stellar mass distributions of local galaxies in addition of an inventory of galaxy mass functions, MFs, for HI, $H_2$, cold gas, and baryonic mass, separately into early- and late-type galaxies. The MFs are determined using the HI and $H_2$ conditional distributions and the galaxy stellar mass function (GSMF). For the conditional distributions we use the results from the compilation presented in Calette et al. [(2018) RMxAA, **54**, 443.]. For determining the GSMF from $M_* \sim 3 \times 10^7$ to $3 \times 10^{12}$ $M_\odot$, we combine two spectroscopic samples from the Sloan Digital Sky Survey at the redshift range $0.0033 < z < 0.2$. We find that the low-mass end slope of the GSMF, after correcting from surface brightness incompleteness, is $\alpha \approx -1.4$, consistent with previous determinations. The obtained HI MFs agree with radio *blind* surveys. Similarly, the $H_2$ MFs are consistent with CO follow-up optically-selected samples. We estimate the impact of systematics due to mass-to-light ratios and find that our MFs are robust against systematic errors. We deconvolve our MFs from random errors to obtain the intrinsic MFs. Using the MFs, we calculate cosmic density parameters of all the baryonic components. Baryons locked inside galaxies represent 5.4% of the universal baryon content, while $\sim$96% of the HI and $H_2$ mass inside galaxies reside in late-type morphologies. Our results imply cosmic depletion times of $H_2$ and total neutral H in late-type galaxies of $\sim$1.3 and 7.2 Gyr, respectively, which shows that late type galaxies are on average inefficient in converting $H_2$ into stars and in transforming HI gas into $H_2$. Our results provide a fully self-consistent empirical description of galaxy demographics in terms of the bivariate gas–stellar mass distribution and their projections, the MFs. This description is ideal to compare and/or to constrain galaxy formation models.

**Keywords:** galaxies: evolution – galaxies: luminosity function – galaxies: mass function

(Received 19 December 2019; revised 28 April 2020; accepted 4 May 2020)

## 1. Introduction

The determination of the matter–energy content of the Universe is one of the most important achievements from the recent advances in observational cosmology (e.g., Planck Collaboration et al. 2016; 2018). Current determinations are fully consistent with the spatially flat $\Lambda$ Cold Dark Matter ($\Lambda$CDM) cosmology, with a present-day matter–energy content dominated by the cosmological constant, $\Omega_\Lambda = 0.689$, and contributions of cold dark matter and baryon matter of $\Omega_{cdm} \approx 0.262$ and $\Omega_{bar} = 0.049$, respectively (for a value of the normalised Hubble constant of $h = 0.674$, Aver et al. 2015; Cooke, Pettini, & Steidel 2018; Planck Collaboration et al. 2018). Therefore, the universal baryon mass density fraction is $f_{bar,U} \equiv \Omega_{bar}/\Omega_m = 0.158$, where $\Omega_m = \Omega_{cdm} + \Omega_{bar}$. How much of these baryons, and their different components, are locked inside galaxies? This paper addresses this question by quantifying the contribution from stars, atomic, and molecular gas in galaxies of different masses and morphological types.

According to the current paradigm of structure formation, non-baryonic dark matter played a major role in the evolution of the non-linear structures that we see today. Particularly, galaxies are believed to form and evolve within extended dark matter haloes, where multiple physical mechanisms are responsible for self-regulating star formation and thus setting up their observed properties (for reviews see, Mo, van den Bosch, & White 2010; Frenk & White 2012; Somerville & Davé 2015). As dark matter structures and galaxies evolve, baryons are redistributed from an initial smooth distribution to a more complex variety of structures. Of primordial importance for galaxy evolution is the amount of neutral hydrogen available for the formation of stars. Gas radiative cooling within the haloes regulates the inflow of cold gas to galaxies. The subsequent formation of stars is regulated by a complex interaction between cold gas inflows and the gas heating/outflows produced by the stars, a process that depends on halo mass. In low-mass halos, the stellar feedback, mostly form Supernova (SN) explosions, is able not only to heat the interstellar medium (ISM) but also to expel large gas fractions from the galaxy. In high-mass haloes, the long cooling time of shock-heated gas and the power-

**Author for correspondence:** Aldo Rodríguez-Puebla, E-mails: apuebla@astro.unam.mx, rodriguez.puebla@gmail.com

ful feedback from rapidly accreting supermassive black holes that heats and/or expels the gas tend to suppress the star formation. Thus, it is not surprising that the expected fraction of baryons inside galaxies will differ from the universal baryon fraction, $f_{bar,U}$. Therefore, constraining the fraction of baryons and their different components in galaxies (mainly stars, atomic, and molecular gas) is essential to constrain the processes that have taken place during the evolution of the galaxies.

One of the main properties of galaxies are their stellar masses $M_*$. Indeed, the abundance of galaxies as a function of $M_*$ provides important clues regarding the evolution of the galaxy population (e.g., Peng et al. 2010, 2012; Yang et al. 2012; Rodríguez-Puebla et al. 2017). Over the last two decades, there has been a remarkable progress in assembling large galaxy samples from multi-wavelength sky surveys that have led to robust determinations of the galaxy stellar mass function (GSMF; for recent discussions, and compilations of observations up to high redshifts see, Conselice et al. 2016; Rodríguez-Puebla et al. 2017). While there have been similar efforts in assembling galaxy samples for atomic gas mass, $M_{HI}$, based on radio blind observations (e.g., Zwaan et al. 2003; Meyer et al. 2004; Koribalski et al. 2004; Kovac et al. 2005; Martin et al. 2010; Haynes et al. 2011; Hoppmann et al. 2015; Haynes et al. 2018) or from follow-up subsamples based on optical/infrared surveys (e.g., Springob et al. 2005; van Driel et al. 2016), these are relatively shallow and/or in small volumes compared to the optical/infrared sky surveys, as well as strongly affected by selection effects. Therefore, the demographical analysis of $M_{HI}$ is challenging especially when determining the low- and high-mass ends of the HI mass function (HI MF) (for a more detailed discussion, see Jones et al. 2018)[a], as well as other statistics like the HI two-point correlation functions[b]. The situation is not that different and even more challenging for the molecular gas as there are not blind galaxy samples in $H_2$. Nonetheless, there are some notable efforts to use optically selected samples combined with small and shallow CO surveys to indirectly derive, from the (uncertain) CO-to-$H_2$ mass conversion factor, the galaxy MF in $H_2$, $H_2$ MF (e.g., Keres, Yun, & Young 2003; Lagos et al. 2014; Saintonge et al. 2017; Andreani et al. 2018). Unfortunately, these CO surveys are also subject to incompleteness and selection effects or subject to a large fraction of galaxies with upper limits reported due to flux detection limits.

As mentioned above, galaxy formation is a non-linear and complex process. Remarkably, well-defined correlations (usually power laws) are, however, found from the observations. Among these are the correlations between the HI and stellar mass, $M_{HI}-M_*$, and the $H_2$ and stellar mass, $M_{H2}-M_*$. While both correlations present large scatters, when divided into early- and late-type galaxies they tend to show different and tighter correlations (e.g., Calette et al. 2018, and more references therein). This is not surprising given that the formation histories of early- and late-type galaxies were different. Thus, understanding the contribution of these two populations to the abundance of galaxies traced by HI and $H_2$ provides further key constrains to galaxy formation theory models.

In a recent work, (Calette et al. 2018, hereafter Paper I) were able to determine empirically not only the mean $M_{HI}-M_*$ and $M_{H_2}-M_*$ relations and their scatters for early- and late-type

galaxies but also the full conditional probability distribution functions (CPDFs) of $M_{HI}$ and $M_{H_2}$ given $M_*$, hereafter HI-CPDF and $H_2$-CPDF, respectively. In this paper, we combine the empirical CPDFs with the $z = 0$ GSMF to derive the bivariate gas-to-stellar mass distributions and the MFs for the HI, $H_2$, cold gas, and baryon components, for all galaxies as well as for early and late types. Thus, the present paper represent a natural continuation of Paper I with some updates. These updates include new constraints on the best-fitting parameters to the observed CPDFs from Paper I.

In this paper, we compute the GSMF and its decomposition into early- and late-type galaxies. While there are many studies that have determined the GSMF in the past, they do not typically report systematic errors or do not deconvolve it from random errors (with a few exceptions, e.g., Bernardi et al. 2010, 2017; Obreschkow et al. 2018) or they are limited in the dynamical range of $M_*$ due to the limited depth of the sample and/or the cosmic variance in the galaxy sample (but see Wright et al. 2017). We combine here two large galaxy samples, the low-redshift sample, low-z, from the NYU Sloan Digital Sky Survey (SDSS) DR4 (Blanton et al. 2005a; b), and the new photometry pipeline for the SDSS DR7 from Meert et al. (2015, 2016). The low-z sample suffers from surface brightness incompleteness, but here we estimate and correct for the fraction of missing galaxies due to this selection effect. As for the SDSS DR7, the new photometry from Meert et al. (2015) shows that galaxy magnitudes were previously underestimated due to sky subtraction problems (see also, Simard et al. 2011); the impact of these new determinations has been studied previously in Bernardi et al. (2017). We extend the Bernardi et al. (2017) analysis using not only different definitions of galaxy stellar masses but by dividing into two morphology groups, early- and late-type galaxies.

The results reported in this paper integrate the HI- and $H_2$-CPDFs with new determinations for the GSMF to offer a full statistical description of the local galaxy demographics traced by the stellar, HI, $H_2$, total cold gas, and baryon mass components. This statistical description of the local galaxy demographics is much more complete than the typically employed GSMF for constraining models and simulations of galaxy formation. The new generation of semi-analytic models (e.g., Croton et al. 2016; Lagos et al. 2018; Henriques et al. 2019; Yung et al. 2019) and cosmological hydrodynamics simulations (e.g., Hirschmann et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Pillepich et al. 2018; Davé et al. 2019), and their post-processing outcomes, are now able to predict stellar, HI, and $H_2$ masses for large galaxy populations in cosmological boxes (see e.g., Lagos et al. 2015; Diemer et al. 2018, 2019; Popping et al. 2019). The empirically based results presented here are optimal for comparing with these predictions as well as for calibrating theoretical models of galaxy evolution (see e.g. Romeo 2020). The results to be presented in this paper are the basis for further studies as the inference of the galaxy–halo connection extended to HI, $H_2$, cold gas, and baryon masses.

The present paper is organised as follows. In Section 2, we describe our method of using CPDFs in order to derive galaxy MFs traced by atomic, molecular, and cold gas masses as well as by the baryonic mass. In Section 3, we describe the samples we use to derive our local GSMF divided into early- and late-type galaxies. In Section 4, we present the results for our inventory of galaxy MFs and compare them with direct observational results. We also present our estimates for the cosmic density parameters related to the different baryonic components in galaxies. Section 5 discusses

---

[a] As we will discuss in Section 4, studying the very low-mass end of the HI MF is beyond the scope of this paper.

[b] Two point correlation functions will be discussed in Calatte et al. in prep.

the impact of systematics and random errors. In Section 6, we present a summary and our main conclusions.

In this paper, we adopt cosmological parameter values that are close to the Planck mission: $\Omega_\Lambda = 0.693$, $\Omega_m = 0.307$, $\Omega_{bar} = 0.048$, and $h = 0.678$. All stellar masses are normalised to a Chabrier (2003) initial mass function (IMF).

## 2. Modelling the bivariate distributions and MFS from the conditional distribution functions

In this section, we describe the statistical method for deriving the HI and $H_2$ MFs (as well as the total cold gas and baryon MFs), from the GSMF and the respective correlations of $M_{HI}$ and $M_{H_2}$ with $M_*$, or more generally, the respective full mass conditional distribution functions, CPDFs. In general, our approach allows to calculate bivariate distribution functions of the HI or $H_2$ mass and the stellar mass. One can imagine that our methodology is equivalent to an optically selected volume-limited sample that it is complete in stellar mass, with HI and $H_2$ gas masses determined for every galaxy in the sample, and for which any MF can be determined. When information about morphology is available, the CPDFs are useful for deriving the corresponding MFs into different morphological components. Here, we consider that the galaxy population is divided into two main morphological groups: early- and late-type galaxies. Following Paper I, our definition of early-type galaxies includes morphological types that comprise E and S0 galaxies or equivalently $T \leq 0$ from the Nair & Abraham (2010) morphology classification. Late-type galaxies are just the complement, from Sa to Irr. Below, we briefly describe the basic ingredients for calculating the MFs:

- **Conditional Distribution Functions**: For a fixed morphology, a galaxy of mass $M_*$ has the chance of having either a HI or $H_2$ mass described by their corresponding CPDFs. We denote the CPDFs of early and late types by $P_E(M_j|M_*)$ and $P_L(M_j|M_*)$, respectively, where $j =$ HI or $H_2$. The HI-CPDF and H2-CPDF contain information about all the moments of the HI- and $H_2$-to-stellar mass correlations. We use the observed HI-CPDF and H2-CPDF from Paper I. In Section 2.2, we describe the functional forms for the CPDFs proposed in Paper I.

- **Galaxy Stellar Mass Function**: The GSMF is an important input since it allows us to project the CPDFs into their corresponding MFs. We derive the GSMF for all galaxies, as well for the early and late type, based on the SDSS. Section 3 describes our methodology to compute the observed GSMF over ∼5 decades in $M_*$, as well as its decomposition into early- and late-type galaxies.

The reader interested in our resulting MFs and bivariate distributions may skip to Section 4.

### 2.1. Generalities

As discussed above, a CPDF, $P_j(M_j|M_*)$, determines the chances that a galaxy of mass $M_*$ possess a specific galaxy property $M_j$, with $j =$ HI, $H_2$, cold gas, or baryonic mass. Note that the units of $P_j$ is per $M_\odot$. The relation between the distribution $P_j$ in bins per $M_\odot$ to dex-1, $\mathcal{P}_j$, is given by

$$\mathcal{P}_j(M_j|M_*) = P_j(M_j|M_*) \times \frac{M_j}{\log e}. \qquad (1)$$

The advantage of using $P_j(M_j|M_*)$ is that it contains information about *all the moments* of the distribution, in particular the mean $M_j - M_*$ relation and its standard deviation.

The *joint* distribution function of $M_*$ and $M_j$, hereafter referred as the *bivariate* distribution function, is defined as:

$$\Phi(M_j, M_*) = \frac{d^2N(M_j|M_*)}{Vd\log M_j d\log M_*} = \mathcal{P}_j(M_j|M_*)\phi_*(M_*), \qquad (2)$$

where $d^2N$ is the bivariate number of galaxies within the mass range $\log M_* \pm d\log M_*/2$ and $\log M_j \pm d\log M_j/2$ in a given volume $V$, and $\phi_*(M_*)$ is the GSMF in units of $Mpc^{-3}dex^{-1}$. The integration (marginalisation) of $\Phi(M_j, M_*)$ over $M_*$ results in the total MF for $M_j$, $\phi_j(M_j)$, that is,

$$\phi_j(M_j) = \int \Phi(M_j, M_*)d\log M_*$$
$$= \int \mathcal{P}_j(M_j|M_*)\phi_*(M_*)d\log M_*. \qquad (3)$$

The above equation shows how the CPDFs are projected into a number density function via the GSMF. Note that integration of $\Phi(M_j, M_*)$ over $M_j$ gives the total GSMF[c].

As discussed previously, when studying the properties of galaxies, it is useful to separate them into, at least, two morphological components such as early types, or spheroid-dominated galaxies, and late types, or disc-dominated galaxies. Thus, the total GSMF can be formally represented as the contribution of these two types

$$\phi_*(M_*) = \phi_{*,E}(M_*) + \phi_{*,L}(M_*), \qquad (4)$$

denoted respectively by $\phi_{*,E}$, and $\phi_{*,L}$. In terms of the fraction of early- and late-type galaxies ($f_E$ and $f_L$), their corresponding galaxy stellar MFs are given respectively by $\phi_{*,E} = f_E \times \phi_*$, and $\phi_{*,L} = f_L \times \phi_*$, with $f_E + f_L = 1$.

Early- and late-type galaxies are different in their HI- and $H_2$-to-stellar mass distributions. Thereby, Equation (3) can be generalised in terms of the distribution $\mathcal{P}_{i,j}(M_j|M_*)$, where the subscripts indicate $i =$ early or late type, and $j =$ HI, $H_2$, cold gas, or baryonic mass. Then, the generalisation of Equation (3) to galaxies with morphological type $i$ and mass component $j$ is

$$\phi_{j,i}(M_j) = \int f_i(M_*)\mathcal{P}_{i,j}(M_j|M_*)\phi_*(M_*)d\log M_*. \qquad (5)$$

Finally, the *total* CPDFs are calculated from the respective conditional distributions of early- and late-type galaxies as:

$$\mathcal{P}_j(M_j|M_*) = f_E(M_*) \times \mathcal{P}_{E,j}(M_j|M_*)$$
$$+ f_L(M_*) \times \mathcal{P}_{L,j}(M_J|M_*), \qquad (6)$$

with $j =$ HI, $H_2$, cold gas, or baryonic mass.

### 2.2. The HI and $H_2$ conditional distribution functions

As shown in Equations (2) and (5), the conditional or bivariate distribution functions are useful to statistically determine the MFs. Evidently, in the case of atomic and molecular gas, we are assuming that for every galaxy that is optically selected, there must exist

---

[c] In the literature there are different methods to determine multivariate joint distributions, one example is the copula approach. A copula is function that join multivariate cumulative distribution functions to their corresponding marginal distributions. They are useful to model the dependence between random variables based on uniform marginals. According to the Sklar's theorem, any multivariate joint distribution is totally defined given the marginal distributions and a copula describing the structure between random variables. For more details on the copula approach and the application to the galaxy luminosity function, the reader is referred to Takeuchi (2010) and Takeuchi et al. (2013). Here, we use the CPDFs formalism for two reasons: (1) the input data that we use are characterised on that format, see Paper I and below; and (2) our goal is to determine the MFs using the CPDFs.

HI and $H_2$ counterparts. The discussion on the possible existence of pure HI or $H_2$ galaxies, those that will not be observed in optically selected samples but rather in radio blind surveys, is out of the scope of this paper. Note that if they exist, the chance of observing those galaxies is very low over the mass ranges that we will derive the MFs. For example, in the case of pure HI galaxies, the Arecibo Legacy Fast ALFA (ALFALFA) survey has found $\sim$1.5% of HI sources that were not clearly associated with an optical counterpart. Of those, $\sim$75% are likely tidal in origin (Haynes et al. 2011). Thus, $\sim$0.4% of HI source observed in the ALFALFA survey are purely gaseous galaxies candidates, most of them at the mass range $10^7 < M_{HI}/M_\odot < 10^{10}$ (Cannon et al. 2015). As we will show, our completeness limit for the HI MF is $M_{HI} \sim 10^8 M_\odot$. The above fraction could be considered as an upper limit as some of these sources have already detected optical counterparts revealing unusual high HI mass-to-light ratios (Cannon et al. 2015). Thus, we conclude that our results are unlikely to be affected by excluding pure gas galaxies in our analysis.

### 2.2.1. The Calette et al. (2018) HI and $H_2$ conditional distribution functions

Here, we use the results from Paper I (Calette et al. 2018) who determined the HI- and $H_2$-to-stellar mass ratio distributions (CPDFs) as a function of $M_*$ from a large compilation of optically selected samples with radio observations. Next, we briefly describe the steps taken in Paper I to derive the HI and $H_2$ CPDFs. The reader is referred to that paper for details.

The compiled data described in Paper I consist of a set of incomplete and inhomogeneous samples. We first homogenised all these samples to a common IMF, cosmology, radio telescope configuration and sensitivity, and CO-to-luminosity conversion factor. Then, we selected only those samples without obvious biases due to selection effects such as environment. Radio non-detections, reported in the literature as upper limits, are an important source of uncertainty when deriving distributions or correlations. In Paper I, we *included* non-detections to derive the HI and $H_2$ CPDFs. Below, we briefly describe the treatment that we employed for radio non-detections.

In our compiled samples, most of radio non-detections are early-type galaxies representing a non-negligible fraction of intermediate and massive galaxies, which are (typically) gas poor. An important fraction of those galaxies are from the GASS (Catinella et al. 2013) and the CO Legacy Database for the GALEX Arecibo SDSS survey (COLD-GASS) (Saintonge et al. 2011) surveys at distance of $109 < D/Mpc < 222$. Compared to other more nearby samples of intermediate and massive early-type galaxies with measurements of HI and $H_2$ mass, such as the ATLAS 3D (Serra et al. 2012) at $\bar{D} \sim 25$ Mpc, we noted that the upper limits of the GASS/COLD-GASS samples are $\sim 1-2$ orders of magnitude larger than nearby samples (Paper I). The above lead us to first introduce a correction for the upper limits of the GASS/COLD-GASS surveys by a distance effect. Recall that radio non-detections or upper limits depend not only on the sensitivity of the radio telescope or integration time but also on the distance to the object. In Paper I, we corrected the upper limits of the GASS/COLD-GASS samples by a distance effect using nearby samples such as the ATLAS 3D survey. Briefly, our correction consists in using the distances and upper limits from nearby samples to estimate the upper limits in the GASS and COLD-GASS as if these two samples were at the same distance as the nearby ones. We validated our procedure using a mock galaxy survey by applying similar distance

sensitivity effects as GASS/COLD-GASS surveys, for details see Paper I. For late-type galaxies, notice that most of them are detected in radio due to their large fractions of gas and it is not necessary to introduce the above corrections. Next, we describe the treatment of the upper limits to derive the HI and $H_2$ CPDFs.

In our analysis from Paper I, we *included* upper limits, or left-censored data, using the Kaplan & Meier (1958) non-parametric estimator. This estimator provides a reconstruction of information lost by censoring. Feigelson & Nelson (1985) adapted this estimator for astronomical samples. We used the Astronomical SURVival statistics (ASURV) package based on Feigelson & Nelson (1985) to derive the HI and $H_2$ CPDFs from our compiled samples. We have also applied the censoring Buckley & James (1979) regression method to derive the relationship and standard deviations between the HI- and $H_2$-to-stellar mass ratio and $M_*$. We note that the regression results are consistent with the (logarithmic) mean and standard deviation values obtained from the CPDFs based on the Kaplan & Meier (1958) estimator.

### 2.2.2. The functional forms of the HI and $H_2$ conditional distribution functions

For the HI and $H_2$ CPDFs of late-type galaxies, in Paper I, we found that they are described by a Schechter function. In the case of early-type galaxies, the CPDFs are better described by a (broken) Schechter function plus a uniform distribution at the low$-\mathcal{R}_j$ values. Following, we describe in more detail these functional forms.

We begin by introducing the following Schechter-type probability distribution function for the HI- or $H_2$-to-stellar mass ratios, $\mathcal{R}_j = M_j/M_*$, in the range $\log \mathcal{R}_j \pm d \log \mathcal{R}_j/2$:

$$\mathcal{S}_{i,j}(\mathcal{R}_j) = \frac{\ln(10)}{\mathcal{N}_{i,j}} \left(\frac{\mathcal{R}_j}{\mathcal{R}_{i,j}^*}\right)^{\alpha_{i,j}+1} \exp\left(-\frac{\mathcal{R}_j}{\mathcal{R}_{i,j}^*}\right), \qquad (7)$$

where the morphology is represented with $i =$ early or late type, and the galaxy property is represented with $j =$ HI or $H_2$. The parameters are the characteristic gas-to-stellar mass ratio, $\mathcal{R}_{i,j}^*$, the normalisation parameter, $\mathcal{N}_{i,j}$, which constrains the probability to be between zero and one,[d] and the power law slope $\alpha_{i,j}$ for the part of the distribution of galaxies with low gas-to-stellar mass ratio.

• Late-type galaxies:

For late-type galaxies, that is $i = L$, in Paper I, we found that the HI-CPDF and H2-CPDF are described by the Schechter-type distribution function given by (7) with the parameters $\alpha_{L,j}$ and $\mathcal{R}_{L,j}^*$ functions of $M_*$ as follows:

$$\alpha_{L,j} = \alpha_{0;L,j} \log M_* + \alpha_{1;L,j}, \qquad (8)$$

and

$$\mathcal{R}_{L,j}^* = \frac{\mathcal{R}_{0;L,j}^*}{\left(\frac{M_*}{\mathcal{M}_{L,j}^*}\right)^{\beta_{L,j}} + \left(\frac{M_*}{\mathcal{M}_{L,j}^*}\right)^{\gamma_{L,j}}}. \qquad (9)$$

Consider that $\mathcal{S}_{L,j}(\log \mathcal{R}_j)d \log \mathcal{R}_j = \mathcal{S}_{L,j}(\log M_j - \log M_*)d \log(M_j/M_*)$. By definition $M_*$ is fixed, thus the HI and $H_2$ CPDFs of late-type galaxies are given by:

$$\mathcal{P}_{L,j}(M_j|M_*)d \log M_j = \mathcal{S}_{L,j}(\log M_j - \log M_*)d \log M_j. \qquad (10)$$

---

[d]For $\alpha_{i,j} > -1$ then $\mathcal{N}_{i,j} = \Gamma(1+\alpha_{i,j})$, with $\Gamma(x)$ the complete gamma function. In general, $\mathcal{N} \propto \int_{-\infty}^{\infty} x^\alpha \exp(-x)dx$.

The above explicitly shows that the integration over conditional distribution functions can also be interpret as convolutions in Equation (3).

- Early-type galaxies:

In the case of early-type galaxies, $i = E$, we showed in Paper I that both for the HI-CPDF and H2-CPDF are described as the sum of two distribution functions; the Schechter-type distribution function, $\mathcal{S}_{E,j}$, and a uniform function, $\mathcal{U}_{0,j}$, are

$$
\mathcal{E}_j(\mathcal{R}_j) = \begin{cases} \mathcal{U}_{0,j} & \mathcal{R}_{0,j} \leq \mathcal{R}_j < \mathcal{R}_{1,j} \\ A \times \mathcal{S}_{E,j}(\mathcal{R}_j) & \mathcal{R}_{1,j} \leq \mathcal{R}_j \end{cases}, \qquad (11)
$$

where $\mathcal{R}_{0,j} = \mathcal{R}_{1,j}/10,$[e] and $\log \mathcal{R}_{1,j} = r_{0,j} \log M_* + r_{1,j}$, while the uniform distribution is given by

$$
\mathcal{U}_{0,j}(M_*) = \frac{p_{0,j} \log M_* + p_{1,j}}{\Delta}, \qquad (12)
$$

and

$$
A = \left(1 - \mathcal{U}_{0,j} \times \Delta\right) \times \frac{\mathcal{N}_{i,j}}{\eta_{i,j}(\mathcal{R}_{1,j})}, \qquad (13)
$$

where in Paper I we assumed that $\Delta = \log_{10} = 1$ dex, the symbol $\eta_{i,j}(\mathcal{R}_{1,j})$ takes into account the fraction of galaxies in the Schechter-type mode for galaxies with gas ratio above $\mathcal{R}_{1,j}$.[f] The HI-CPDF and $H_2$-CPDF of early-type galaxies are

$$
\mathcal{P}_{E,j}(M_j|M_*)d \log M_j = \mathcal{E}_j(\log M_j - \log M_*)d \log M_j. \qquad (14)
$$

### 2.2.3. Constraints on the best-fitting parameters

In Paper I, the best-fit parameters for late-type galaxies, Equations (8)–(9), and for early-type galaxies, Equations (11)–(13), were constrained using the observed HI- and $H_2$-CPDFs on various stellar mass bins. Computing CPDFs over $M_*$ bins requires the GSMF in addition of the fraction of early- and late-type galaxies (see Section 4.1 for more details). Since we are using slightly different inputs, namely the GSMF and the fractions of early- and late-type galaxies, than in Paper I, we prefer to perform our own fits to the same data, for consistency. The results are presented in Section 4.1. The differences with the parameters reported in Paper I are actually small.

### 2.3. The cold gas and baryonic conditional distribution functions

Once we have constructed the HI-CPDF and $H_2$-CPDF, we can now define the conditional distributions for the cold gas and baryon masses, $M_{gas}$ and $M_{bar}$.

The total cold gas content in a galaxy is composed of HI, $H_2$, helium, and metals; helium and metals account for roughly 30% of the cold gas, $M_{He} + M_Z \approx 0.3 M_{gas}$ Therefore, $M_{gas} = M_{HI} + M_{H_2} + M_{He} + M_Z = 1.4 \times (M_{HI} + M_{H_2})$. For simplicity, let $M_{HI}$ and $M_{H_2}$ be two independent random variables. Section 5 discusses

the validity of this assumption. Then, $M_{gas}$ is a random variable with the conditional distribution function:

$$
\begin{aligned}
P_{gas}(M_{gas}|M_*) &= \frac{1}{1.4} \int P_{HI}\left(0.71 M_{gas} - M_{H_2}|M_*\right) \\
&\quad \times P_{H_2}(M_{H_2}|M_*)dM_{H_2}, \\
&= \frac{1}{1.4} \int P_{HI}(M_{HI}|M_*) \\
&\quad \times P_{H_2}(0.71 M_{gas} - M_{HI}|M_*)dM_{HI}, \qquad (15)
\end{aligned}
$$

or after some algebra, the same distribution function but per bin in log space is:

$$
\begin{aligned}
\mathcal{P}_{gas}(M_{gas}|M_*) &= \int \frac{\mathcal{P}_{HI}(0.71 M_{gas} - M_{H_2}|M_*)}{1 - 1.4\, M_{H_2}/M_{gas}} \\
&\quad \times \mathcal{P}_{H_2}(M_{H_2}|M_*)d \log M_{H_2}, \\
&= \int \mathcal{P}_{HI}(M_{HI}|M_*) \\
&\quad \times \frac{\mathcal{P}_{H_2}(0.71 M_{gas} - M_{HI}|M_*)}{1 - 1.4\, M_{HI}/M_{gas}}d \log M_{HI}. \qquad (16)
\end{aligned}
$$

For the baryonic conditional distribution functions, we again assume that $M_{gas}$ and $M_*$ are two independent random variables. Thus, $M_{bar} = M_{gas} + M_*$ is a random variable with a distribution function given by

$$
\begin{aligned}
P_{bar}(M_{bar}|M_*) &= \int P_{gas}(M_{bar} - \mathcal{M}_*|M_*) \\
&\quad \times \delta(\mathcal{M}_* - M_*)d\mathcal{M}_* \qquad (17) \\
&= P_{gas}(M_{bar} - M_*|M_*),
\end{aligned}
$$

where $P_{gas}$ is the conditional distribution function for gas, Equation (15), and the Dirac-$\delta$ function appears explicitly for the $M_*$ term. Similarly as above, we find that

$$
\mathcal{P}_{bar}(M_{bar}|M_*) = \frac{\mathcal{P}_{gas}(M_{bar} - M_*|M_*)}{1 - M_*/M_{bar}}. \qquad (18)
$$

Finally, we derive the gas and baryon MFs using Equations (5), (16), and (18), the last two valid for early- and late-type galaxies.

## 3. The GSMF of all, early-, and late-type galaxies

The preceding section described a methodology to use the GSMF as an interphase that transforms galaxy CPDFs into MFs, see Equation (5). In this section, we briefly describe how we determine the local GSMF for masses above $M_* \sim 3 \times 10^7 M_\odot$, as well as the GSMF s for early- and late-type galaxies. For a more detailed description of the galaxy samples utilised here and the different corrections we apply, the reader is referred to Appendices A–C.

### 3.1. The galaxy samples and the GSMF

To estimate the GSMF over a large dynamical range, we use two galaxy samples. Next, we shortly describe the procedure and our determinations.

(1) For masses above $M_* = 10^9 \, M_\odot$, we use the SDSS DR7 based on the photometric catalogue from Meert et al. (2015) and Meert et al. (2016)[g] at the redshift interval $0.005 < z < 0.2$. Previous studies have concluded that the measurements of the apparent

---

[e]As discussed in Paper I, the observed data imply that the HI- and $H_2$-to-$M_*$ ratios will not be lower than $\sim 10^{-4} - 10^{-5}$. This seems plausible since even for galaxies that transformed all their gas into stars, the gas mass recycled to the ISM by stellar evolution could provide the above minimal floor for the gas mass ratios.

[f]Similarly to late types, in the case that $\alpha_{i,j} > -1$, then $\eta_{i,j}(\mathcal{R}_{1,j}) = \gamma(1 + \alpha_{i,j}, \mathcal{R}_{1,j})$, with $\gamma(x, a)$ as the incomplete gamma function. In general, $\eta_{i,j}(a) \propto \int_a^\infty x^\alpha \exp(-x)dx$.

[g]Available at http://www.physics.upenn.edu/?ameert/SDSS_PhotDec/

brightnesses based on the standard SDSS pipeline photometry are underestimated due to sky subtraction problems, particularly, in crowded fields (Bernardi et al. 2010; Blanton et al. 2011; Simard et al. 2011; Bernardi et al. 2013; He et al. 2013; Mendel et al. 2014; Kravtsov, Vikhlinin, & Meshcheryakov 2014; Meert et al. 2015; D'Souza, Vegetti, & Kauffmann 2015; Bernardi et al. 2016; Meert et al. 2016. New determinations of the GSMF based on the new algorithms for obtaining more precise measurements of the sky subtraction, and thus to improve the photometry, have concluded that the bright end of the luminosity/MF has been systematically underestimated (Bernardi et al. 2017). While there are various groups working in improving the determination of galaxy apparent brightnesses, see references above, Bernardi et al. (2017) showed that all those studies agreed up to 0.1 dex in the GSMF. In this paper, we use the apparent Sérsic *r*, *g*, and *i* band luminosities reported in Meert et al. (2015) and (2016) derived for the SDSS DR7 based on the PyMorph software pipeline (Vikram et al. 2010; Meert, Vikram, & Bernardi 2013). This software has been extensively tested in Meert et al. (2013) and shows that it does not suffer from sky subtraction problems. All magnitudes and colours are K+E corrected at a redshift rest frame $z = 0$, see Appendixes A and B. As described in Appendix A, for every galaxy, we estimate $M_*$ from five colour-dependent mass-to-light ratios but we define as our fiducial $M_*$ the geometric mean of all the determinations. Using the $1/V_{max}$ method, we derive six GSMFs based on the mass definitions described above. Consistent with Bernardi et al. (2017), we find that the differences in mass-to-light ratios introduce large discrepancies in the GSMF, especially at the high-mass end. In Figure 11 from Appendix A, we find that a shift of $\sim \pm 0.15$ dex in the $M_*$ axis recovers systematic errors in the GSMF due to different mass-to-light ratios.

(2) For masses below $M_* = 10^9\,M_\odot$, we use the SDSS DR4 NUY-VAGC low-z sample,[h] at the redshift interval $0.0033 < z < 0.005$, and ideal to study the low mass/luminosity galaxies (Blanton et al. 2005a; Blanton et al. 2005b). As before, all absolute magnitudes and colours were K+E corrected at a redshift rest frame $z = 0$. Also, we derive $M_*$ from five colour-dependent mass-to-light ratios and, again, we define our fiducial $M_*$ as the geometric mean of all the determinations. We construct the GSMF using the $1/V_{max}$ method and include missing galaxies due to surface brightness incompleteness, as described in Appendix C. For surface brightness incompleteness, we follow closely the methodology described in Blanton et al. (2005b). The latter correction is relevant for the low-mass end. Based on the conclusions from Baldry et al. (2012), we use a simple correction for the low-mass end in order to correct for the local flow model distances from Willick et al. (1997) to the one by Tonry et al. (2000).

Our final GSMF is the result of combining the SDSS NUY-VAGC low-z sample, for galaxies with masses $M_* \approx 3 \times 10^7\,M_\odot$ to $M_* \approx 10^9\,M_\odot$, and the SDSS DR7 sample for galaxies with $M_* \gtrsim 10^9\,M_\odot$, based on our fiducial $M_*$ determination. Figure 1 presents our final GSMF with the black solid circles and error bars. The black solid line shows the best fit to the data (described below), and the grey shaded area shows a shift in the $M_*$ axis of $\pm 0.15$ dex. As discussed above, in Appendix A, we find that this is a good approximation to the systematic errors in the GSMF due to differences in the mass-to-light ratios. In the same figure, we include comparisons to previous works. In order to account for differences
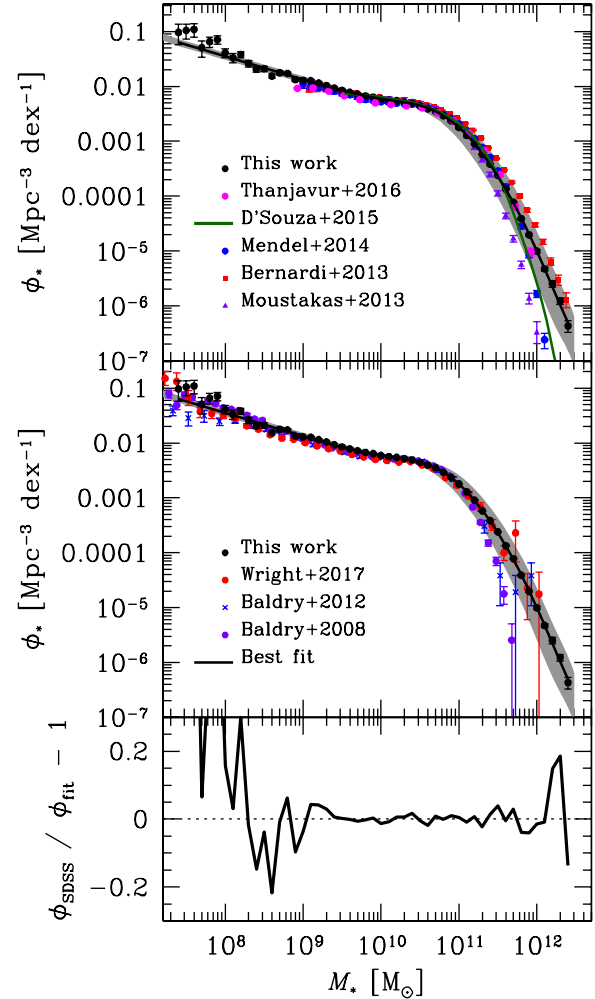
**Figure 1.** Observed GSMF when combining the SDSS NYU-VAGC low-redshift sample and the SDSS DR7 sample, black filled circles with error bars. We reproduce our results in the upper and the middle panels. The best-fit model composed of a Schechter function with a sub-exponential slope and a double power law function is shown as the black solid line. The shaded area shows an estimate of the systematic errors with respect to the best-fitting model. The bottom panel shows the residuals for our best-fitting model as a function of $M_*$. We include comparisons to some previous observational determinations of the GSMF: in the upper panel we show determinations that are complete down to $\sim 10^9\,M_\odot$, mostly based on the SDSS DR7, while in the middle panel we show determinations based on the GAMA survey, which are complete down to $\sim 3-5 \times 10^7\,M_\odot$, but suffer from cosmic variance at high masses due to the small volume.

in cosmologies, we scale previous studies to our cosmology using the following relations:

$$\phi_{*,us} = \phi_{*,lit} \left( \frac{h_{us}}{h_{lit}} \right)^3 , \qquad (19)$$

and

$$M_{*,us} = M_{*,lit} \left( \frac{h_{lit}}{h_{us}} \right)^2 , \qquad (20)$$

where $h_{us} = 0.678$ and $h_{lit}$ is the respective value reported in the literature. Nonetheless, the impact of accounting for different cosmologies is small.

In the upper panel of Figure 1, we reproduce the GSMFs from previous determinations with stellar mass completeness above

$M_* \sim 10^9$ M$_\odot$. The violet triangles with error bars are the determinations from Moustakas et al. (2013), who used a spectroscopic sample of SDSS DR7 galaxies from the New York University Value Added Galaxy Catalog (NYU-VAGC) with redshifts $0.01 < z < 0.2$ combined with observations from GALEX. The red squares with error bars are the estimation obtained in Bernardi et al. (2013) from a sample of SDSS DR7 galaxies with photometry based on the PYMORPH software pipeline at $z \sim 0.1$. Here, we reproduce their result based on Sérsic luminosities. Additionally, we compute the GSMF using the stellar mass estimates from Sérsic photometry from Mendel et al. (2014) who used the Simard et al. (2011) SDSS DR7 sample of $g$ and $r$ band photometry and extended to $u$, $i$, and $z$ bands, blue filled circles with error bars. We show the best-fitting model from D'Souza et al. (2015), who estimated the GSMF by stacking images of galaxies with similar stellar masses and concentrations to correct MODEL magnitudes from the SDSS DR7, dark green solid line. Finally, we compare our result to Thanjavur et al. (2016), who derived the GSMF using the analysis from Mendel et al. (2014).

Our GSMF agrees well with previous determinations at the $\sim 10^{9.3} - 10^{11}$ M$_\odot$ range. At the high-mass end, it is shallower than previous determinations (e.g., Moustakas et al. 2013) except to Bernardi et al. (2013), who use Sérsic photometry from the SDSS DR7. As extensively discussed in Bernardi et al. (2017), there are two systematic effects that could lead to differences when comparing to previous determinations from the literature; assumptions on mass-to-light ratios and estimations of galaxy surface brightness. In the case of Moustakas et al. (2013) and D'Souza et al. (2015), who used CMODEL and MODEL magnitudes, the comparison is not obvious due to systematic effects in both mass-to-light ratios and photometry (Bernardi et al. 2017). In the case of Mendel et al. (2014) and Bernardi et al. (2013), effects on photometry are not the dominant ones but mass-to-light ratios. Nonetheless, those differences are within the expected systematic effect, especially at the massive end, (Bernardi et al. 2017, see also Figure 11). We therefore conclude that when comparing to other previous determinations, the differences that we observe are consistent with the differences expected from systematic effects. Indeed, Figure 1 shows that most of the previous determinations are within our region of systematic errors. Thus, hereafter we will assume that our shift of $\pm0.15$ dex in the $M_*$ axis approximately captures systematics not only from stellar population models but also from photometry.

The middle panel of Figure 1 presents comparisons to some previous determinations from deep but small-volume samples. The purple dots with error bars are from Baldry et al. (2008), who used the SDSS NYU-VAGC low-z sample but did not include missing galaxies due to surface brightness incompleteness. In addition, we compare to Baldry et al. (2012), who used the Galaxy And Mass Assembly (GAMA) survey for galaxies at $z < 0.06$, and complete down to $r = 19.4$ mag for two-thirds of the galaxy sample and to $r = 19.8$ for one-third of the sample. Finally, we reproduce the observed GSMF from Wright et al. (2017), who also used the GAMA survey to estimate the GSMF.

At low masses, our results are in excellent agreement with the GAMA GSMFs. This is encouraging since the GAMA survey does not suffer from surface brightness incompleteness, at least within the stellar mass range that we are comparing our results. This is an indication that the surface brightness corrections described in Appendix C are able to recover the slope of the GSMF at low masses. Consistent with the values reported in Baldry et al. (2012)

and Wright et al. (2017), we find that the faint-end slope of the GSMF is $\alpha \approx -1.4$, below we describe in more detail the fitting model for the GSMF. The above is also in good agreement with Sedgwick et al. (2019) who recently determined the low-mass end of the GSMF by identifying low surface brightness galaxies based on data of core-collapse supernovae. The authors used the IAC Stripe 82 legacy project (Fliri & Trujillo 2016) and the SDSS-II Supernovae Survey (Frieman et al. 2008).

At the massive end we notice, however, some apparent tension between our and the GAMA results. Effects due to cosmic variance (due to the small redshift and angular coverage of the GAMA sample) could explain those differences as well as systematics in the mass-to-light ratios. Indeed, we see that some of the data are within the systematic errors. In addition, note that Figure C5 from Appendix C shows that using the mass-to-light ratios from Taylor et al. (2011), utilised in the Baldry et al. (2012) GSMF, tend to underestimate the high-mass end of the GSMF.

### 3.2. Best-fitting model to the GSMF

To provide an analytic form to our GSMF, we choose to use a function composed of a Schechter function with a sub-exponential decreasing slope and a double power law function. Note that the resulting high-mass end of our GSMF is shallower than an exponential function, and, thus, better fitted to a power law (see also Tempel et al. 2014). The Schechter sub-exponential function is given by:

$$\phi_{*,S}(M_*) = \phi_S^* \ln 10 \left( \frac{M_*}{\mathcal{M}_S} \right)^{1+\alpha_S} \exp\left[ -\left( \frac{M_*}{\mathcal{M}_S} \right)^\beta \right], \quad (21)$$

where $\phi_S^*$ is the normalisation parameter in units of Mpc$^{-3}$ dex$^{-1}$, $\alpha$ is the slope at the low-mass end, $\mathcal{M}_S$ is the characteristic mass, and $\beta$ is the parameter that controls the slope at the massive end; note that $\beta = 1$ corresponds to a Schechter function. The double power law function is given by:

$$\phi_{*,D}(M_*) = \phi_D^* \ln 10 \left( \frac{M_*}{\mathcal{M}_D} \right)^{1+\alpha_D} \left[ 1 + \left( \frac{M_*}{\mathcal{M}_D} \right)^\gamma \right]^{\frac{\delta - \alpha_D}{\gamma}}, \quad (22)$$

where $\phi_D^*$ is the normalisation parameter in units of Mpc$^{-3}$ dex$^{-1}$, $\alpha$ and $\delta$ control the slope at low and high masses, respectively, while $\gamma$ determines the speed of the transition between the low- and high-mass regimes; and $\mathcal{M}_D$ is the characteristic mass of the transition. Finally, the analytic form for fitting the observed GSMF is given by

$$\phi_{*,\text{model}}(M_*) = \phi_{*,S}(M_*) + \phi_{*,D}(M_*), \quad (23)$$

where we assumed that $\mathcal{M}_S = \mathcal{M}_D$.

We find the best-fit parameters $\vec{p}_{\text{GSMF}} = (\phi_S^*, \alpha_S, \mathcal{M}_S, \beta, \phi_D^*, \alpha_D, \delta, \gamma)$ that maximise the likelihood function $\mathcal{L} \propto \exp(-\chi^2/2)$ using the Markov chain Monte Carlo (MCMC) method algorithm described in Rodríguez-Puebla et al. (2013). Here

$$\chi^2 = \sum_{i=1}^{N_{\text{obs}}} \left( \frac{\phi_{*,\text{SDSS}}^i - \phi_{*,\text{model}}^i}{\sigma_{\text{SDSS}}^i} \right)^2, \quad (24)$$

with $N_{\text{obs}}$ as the number of observational data points of the GSMF each with an $i$th value of $\phi_{*,\text{SDSS}}^i$ and an error of $\sigma_{\text{SDSS}}^i$. The $i$th value of our model is given by $\phi_{*,\text{model}}^i$.

We sample the best-fit parameters by running a set of 10 chains with $1 \times 10^5$ MCMC models each. Table 1 lists the best-fit parameters. For our best-fitting model, we find that $\chi^2 = 85.42$ from

**Table 1.** Best-fitting parameters for the GSMF Equations (21)–(23)

| $\log \phi_S^*$ [Mpc$^{-3}$dex$^{-1}$] | $\alpha_S$ | $\beta$ | $\mathcal{M}_D = \mathcal{M}_S$ [M$_\odot$] | $\log \phi_D^*$ [Mpc$^{-3}$dex$^{-1}$] | $\alpha_D$ | $\delta$ | $\gamma$ |
|---|---|---|---|---|---|---|---|
| $-3.019 \pm 0.067$ | $-1.418 \pm 0.025$ | $0.660 \pm 0.011$ | $10.897 \pm 0.036$ | $-2.267 \pm 0.120$ | $-0.207 \pm 0.169$ | $3.660 \pm 0.347$ | $1.236 \pm 0.080$ |

a number of $N_{\rm obs} = 50$ observational data points. Our model consist of $N_{\rm p} = 8$ free parameters, thus the reduced $\chi^2$ is $\chi^2/{\rm d.o.f.} = 2.03$. The upper and middle panels of Figure 1 show our best-fitting model as the black solid line and the bottom panel shows the residuals as a function of $M_*$. Our best-fitting model has an error of $\sim 2\%$ at the range $M_* \sim 2 \times 10^9 - 5 \times 10^{11} {\rm M}_\odot$ and an error lower than $\sim 10\%$ at the mass range $M_* \sim 7 \times 10^8 - 1 \times 10^{12} {\rm M}_\odot$. For lower masses, errors can be up to $\sim 20\%$.

A valid question is how much we improve the analytic prescription when using a Schechter sub-exponential plus a double power law function model confronted to a double Schechter function model, commonly employed by previous authors (see e.g., Baldry et al. 2012; Wright et al. 2017). We have explored this possibility but assuming a Schechter function, $\beta = 1$ in Equation (21), and a Schechter sub-exponential function, that is, we are adding a extra degree of freedom due to the shallow decay at the high-mass end. Based on this alternative, we repeat our fitting procedure but this time finding that $\chi^2 = 662.817$ from a number of $N_{\rm obs} = 50$ observational data points. Now, our model consists of $N_{\rm p} = 6$ free parameters resulting in a reduced $\chi^2$ of $\chi^2/{\rm d.o.f.} = 15.06$. This is considerably worse when combining Schechter sub-exponential and double power law functions. Thus, hereafter, we will consider only the latter model.

### 3.3. The GSMFs of early- and late-type galaxies

Our main goal for this paper is to construct bivariate distributions as well as MF based on the observed gas mass CPDFs and the GSMFs of early- and late-type galaxies. In this section, we determine the GSMF of early- and late-type galaxies from the SDSS DR7 spectroscopic sample with the public automated morphological galaxy classification by Huertas-Company et al. (2011).[i] The morphological classification in Huertas-Company et al. (2011) was determined based on support vector machine algorithms. Here we use their tabulated probabilities for each SDSS galaxy as being classified as an early type, $P(E)$. For masses below the completeness of the SDSS DR7 sample, we use an extrapolation of the observed fraction of early-type galaxies. We will come back to this point later in this section.

From a catalogue of galaxies with visual morphological classification (UNAM-KIAS; Hernández-Toledo et al. 2010), we find that galaxies with types $T \leq 0$ are mostly those with $P(E) > 0.65$, and those with $P(E) \leq 0.65$ correspond mostly to $T > 0$; here $T$ is the Fukugita et al. (2007) notation.[j] Based on the above, we consider as early-type galaxies those with a probability $P(E) > 0.65$, while late-type galaxies those with $P(E) \leq 0.65$. We checked that our morphology definition between early- and late-type galaxies is consistent with the morphological classification based on

the concentration parameter $c = R_{90}/R_{50}$. That is, the division between early- and late-types is approximately at $c = 2.85$ (see below and also, Hyde & Bernardi 2009; Bernardi et al. 2010).

We calculate the SDSS DR7 GSMF of early- and late-type galaxies using the $1/V_{\rm max}$ method described in Appendix A. Figure 2 shows the corresponding GSMFs of early and late types in the upper left and right panels, respectively. For comparison, we show the GSMFs for red and blue galaxies based on a colour cut limit in the $(g - r)^{0.0} - M_*$ diagram. In this diagram, we find that a rough division criteria from blue to red galaxies is given by the colour limit of $(g - r)^{0.0} = 0.66$.[k] In the same figure, we compare our results to different determinations from the literature as we describe below. All the data in this figure have been renormalised to our cosmology.

Recently, Moffett et al. (2016a) visually classified morphologies in the GAMA survey and reported the GSMF for different morphologies. Here, we reproduce their GSMF from E to Sa galaxies as early types and the complement as late types. Contrary to our definition, Sa galaxies are included in the early-type group; this is because the authors report S0 and Sa galaxies as one morphology group. As shown in Figure 2, the GSMF of early-type galaxies from Moffett et al. (2016a) results in an overabundance of low-mass galaxies compared to other studies. We reproduce the results from Thanjavur et al. (2016) with bulge-to-total ratios of $B/T > 0.8$ as early types and $B/T \leq 0.8$ as late types. Thanjavur et al. (2016) used the bulge-to-disc decomposition from Simard et al. (2011) SDSS DR7 spectroscopic sample, and stellar masses derived from Mendel et al. (2014). We also include results from Kelvin et al. (2014). Similarly to Moffett et al. (2016a), Kelvin et al. (2014) visually classified morphologies in the GAMA survey. We again use their GSMF from E to Sa galaxies for early types since the authors combined S0–Sa galaxies as in Moffett et al. (2016a). The filled triangles with error bars show the GSMF from Bernardi et al. (2010) for galaxies with concentration parameter $c > 2.86$ for early types and $c \leq 2.86$ for late types.[l]

Finally, using the Nair & Abraham (2010) morphology catalogue, who visually classified 14,034 spectroscopic galaxies from the SDSS DR4, we derive the GSMF for early-type galaxies.[m] We utilise their morphological notation and define early-type galaxies as those objects with $-5 \leq T \leq 0$ (E-S0s), equivalent to $T \leq 1$ in the Fukugita et al. (2007) notation. We additionally derive the GSMF with morphologies between $-5 \leq T \leq 1$ in the Nair & Abraham (2010) notation, which include Sa galaxies.

In general, our results agree with previous determinations, especially with those from the SDSS spectroscopic samples.

---

[i] http://gepicom04.obspm.fr/sdss_morphology/Morphology_2010.html

[j] Huertas-Company et al. (2011) define as elliptical galaxies those objects with $T \leq 0$, S0s as $T = 1$, Sabs as $2 < T < 4$, and Scd as $4 \leq T < 7$ based on the Fukugita et al. (2007) morphology classification. Huertas-Company et al. (2011) included elliptical galaxies and S0s as early-type galaxies which corresponds to galaxies with types $T \leq 1$ in the Fukugita et al. (2007) notation, and $T \leq 0$ when using the Nair & Abraham (2010) notation, see below. In the de Vacouleurs notation, this is equivalently to $T = 0$.

[k] While this is just a rough division line, we used it as a practical method for decomposing the GSMF into two main groups. Notice that in Appendix C, we apply a more sophisticated method to derive the distribution of blue and red galaxies. Additionally, we checked that both methods give similar results.
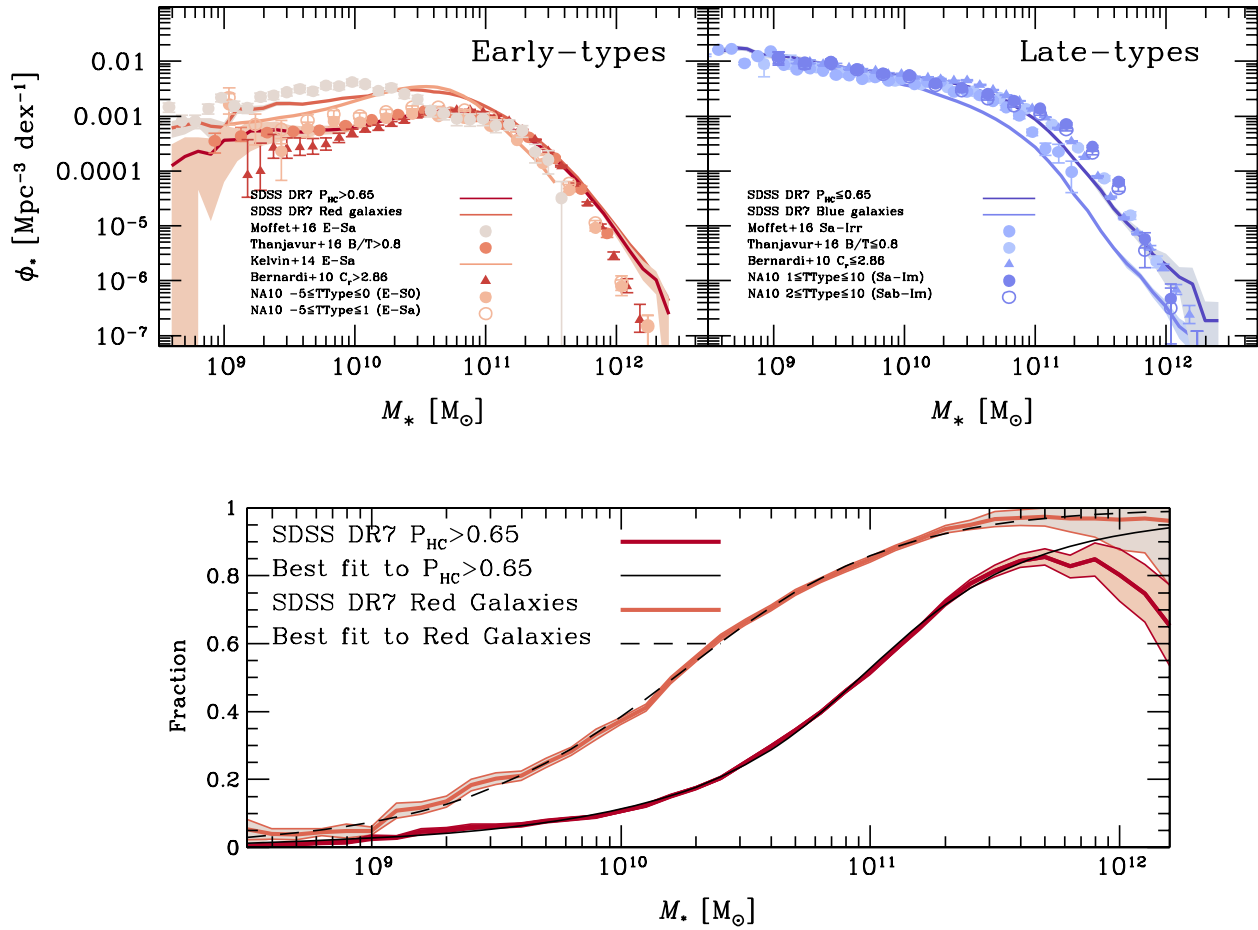
[l] Figure 5 in Bernardi et al. (2010) shows that using $c = 2.85$ separates galaxies into earlier and later morphologies. While this selection criteria is not perfect, their Figure 18 shows that using the above concentration is very similar to the E+S0 GSMF based on the Fukugita et al. (2007) sample.

[m] We construct volume-limited samples that are complete in $M_*$ and compute the GSMF as described in Appendix C. In this case, we slightly modified Equation (C8) by shifting our stellar mass limit by 0.4 dex, that is, $\log M_{*,\rm lim,NA10}(z) = \log M_{*,\rm lim}(z) + 0.4$.

**Figure 2.** SDSS DR7 GSMFs for early- and late-type galaxies, left and right upper panels, respectively. Early- and late-type galaxies are defined as those with $P(E) > 0.65$ ($P(E) \leq 0.65$) from the tabulated probabilities of Huertas-Company et al. (2011). This is equivalent to morphological types that comprises E and S0 galaxies or $T \leq 0$ (Sa to Irr galaxies or $T > 0$). We compare to various previous determinations from the literature as indicated by the legends, see also the text for details. Our determinations are in general in good agreement with previous determinations from SDSS spectroscopic samples, while a tension is evident with determinations from the GAMA survey. We also present our resulting GSMFs for blue and red galaxies. These GSMFs follow closely those by morphology from the GAMA survey. The bottom panel shows our number density-weighted fractions of early-type and red galaxies as a function of $M_*$. Their corresponding best-fit models Equation (25) are shown with solid and dashed lines, respectively.

In contrast, the GSMF of early-type galaxies from the visual classification of the GAMA survey are systematically above our results at the low-mass end, $M_* \lesssim 2 \times 10^{10} M_\odot$, but closer to our classification based on galaxy colour. While it is not clear the reason of the differences outlined above (the inclusion of Sa galaxies as early types, environment, etc.), in Appendix C.4, we will discuss the impact of using galaxy colour instead of morphology when deriving the HI, $H_2$, cold gas, and baryonic MFs separated into two main galaxy populations.

Finally, the bottom panel of Figure 2 shows the resulting fraction of early-type galaxies as a function of stellar mass, $f_E(M_*)$. In addition, we show the fraction of red galaxies when using our $g - r$ colour cut limit, $f_r(M_*)$. We find that the fraction at which early-type galaxies is 50% is at $M_* \sim 10^{11} M_\odot$, while at $M_* \sim 1.6 \times 10^{10} M_\odot$ and $M_* \sim 8 \times 10^{11} M_\odot$ the fractions are 16% and 84%, respectively. For red galaxies, the fraction of 50% is at $M_* \sim 10^{10} M_\odot$, while at $M_* \sim 3 \times 10^9 M_\odot$ and $M_* \sim 10^{11} M_\odot$ the fractions are 16% and 84%, respectively. Note that the characteristic mass at which the fraction of early-type galaxies is 50% is a factor of $\sim 10$ larger than for red galaxies. In general, $f_E(M_*)$ rises slower than the fraction $f_r(M_*)$. In the same figure, we present the best-fit model to the data. After exploring different functions, we find that two

*sigmoid* functions accurately describe the functionality of $f_E(M_*)$ or $f_r(M_*)$:

$$f_k(M_*) = \frac{1-A}{1 + e^{-\gamma_1(x_{C,1}+x_{0,1})}} + \frac{A}{1 + e^{-\gamma_2(x_{C,2}-x_{0,2})}}, \quad (25)$$

where $k = E$ or $r$, $x_{C,i} = M_*/\mathcal{M}_{C,i}$, with $i = 1, 2$. The best-fit parameters for the two fractions are listed in Table 2.

To derive the analytic model for the GSMF of early- and late-type galaxies, we use the best-fit model to our GSMF, Section 3.2, and the best-fit model for $f_E(M_*)$. For masses below $5 \times 10^8 M_\odot$ we extrapolate $f_E(M_*)$. This is an acceptable approximation since as seen in Figure 2, the fraction $f_E(M_*)$ tends to $\sim 0$ below $M_* = 10^9 M_\odot$. Recall that our main goal in this paper is to derive the MFs for HI, $H_2$, cold gas, and baryons by combining the observed HI and $H_2$ CPDFs with the GSMF, both for early- and late-type galaxies, over a large mass range. Thus, at this point, we are in a position to determine these MFs.

## 4. Results

In this section, we present our fits to the HI- and $H_2$-CPDFs for early- and late-type galaxies from Paper I, the corresponding

**Table 2.** Best-fit parameters to the fraction of early-type and red galaxies

| Sample | $A$ | $\gamma_1$ | $\log \mathcal{M}_{C,1}[\mathrm{M}_\odot]$ | $x_{0,1}$ | $\gamma_2$ | $\log \mathcal{M}_{C,2}[\mathrm{M}_\odot]$ | $x_{0,2}$ |
|---|---|---|---|---|---|---|---|
| $P_{HC} > 0.65$ | 0.46 | 3.75 | 11.09 | 0.09 | 1.51 | 10.38 | 0.462 |
| Red galaxies | 0.21 | 2.44 | 10.66 | 0.36 | 1.81 | 9.68 | 0.070 |

**Table 3.** Best-fit parameters of the $HI$ and $H_2$ mass CPDFs for late- and early-type galaxies

| | Late-type galaxies Equations (8)–(10) | | | | | |
|---|---|---|---|---|---|---|
| Component | $\alpha_{0;l,j}$ | $\alpha_{1;l,j}$ | $\mathcal{R}^*_{0;l,j}$ | $\mathcal{M}^*_{l,j}$ | $\beta_{l,j}$ | $\gamma_{l,j}$ |
| HI | $-0.127 \pm 0.036$ | $1.279 \pm 0.345$ | $2.598 \pm 0.745$ | $8.646 \pm 0.399$ | $-0.018 \pm 0.108$ | $0.577 \pm 0.053$ |
| $H_2$ | $-0.085 \pm 0.120$ | $0.830 \pm 1.213$ | $0.122 \pm 0.037$ | $10.595 \pm 0.301$ | $0.841 \pm 0.195$ | $0.063 \pm 0.089$ |

| | Early-type galaxies Equations (11)–(14) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Component | $\alpha_{0;e,j}$ | $\alpha_{1;e,j}$ | $\mathcal{R}^*_{0;e,j}$ | $\mathcal{M}^*_{e,j}$ | $\beta_{e,j}$ | $\gamma_{e,j}$ | $p_{0,j}$ | $p_{1,j}$ | $r_{0,j}$ | $r_{1,j}$ |
| HI | $-0.052 \pm 0.067$ | $-0.074 \pm 0.6840$ | $1.573 \pm 0.533$ | $8.354 \pm 0.258$ | $-0.820 \pm 0.272$ | $0.468 \pm 0.077$ | $0.060 \pm 0.032$ | $-0.113 \pm 0.338$ | $-0.259 \pm 0.015$ | $-0.310 \pm 0.160$ |
| $H_2$ | $0.059 \pm 0.069$ | $-1.491 \pm 0.725$ | $0.674 \pm 0.229$ | $8.182 \pm 0.317$ | $-0.686 \pm 0.412$ | $0.375 \pm 0.156$ | $0.017 \pm 0.074$ | $0.515 \pm 0.785$ | $-1.084 \pm 0.074$ | $7.980 \pm 0.724$ |

correlations (first and second moments), the bivariate HI- and $H_2$-stellar mass distributions, and the HI and $H_2$ MFs. We also present the total cold gas and baryonic MFs. We will show that our empirically inferred HI and $H_2$ MFs agree with direct determinations from *blind* or optically/infrared (selected) radio galaxy surveys. Previous works related to our approach are Obreschkow & Rawlings (2009), Lemonias et al. (2013), and Butcher et al. (2018).

For those interested in using our results, we provide a PYTHON code containing all the necessary information to reproduce the results presented here, for details see Section 6.

### 4.1. The HI *and* H₂ *conditional distribution functions*

Section 2.2 describes the functional forms for the HI- and $H_2$-CPDFs of early- and late-type galaxies proposed in Paper I. Here, using the determinations of the CPDFs for early- and late-type galaxies from Paper I, we find the best-fit parameters of the proposed functional forms: a Schechter-type function and a Schechter-type + Uniform function, respectively (see Section 2.2). While Paper I reported their corresponding best-fit parameters, here we opt for an update based on our own determinations of the GSMFs, for consistency. There are two reasons for doing this: *(i)* when fitting a CPDF that is determined over stellar mass bins, one should take into account contributions to this CPDF from the different masses. Weighting the conditional distributions by the GSMF takes care on that, see Equations (6) and (8) of Section 5 from Paper I; and *(ii)* Paper I used the fraction of bulge-dominated galaxies from Moffett et al. (2016b) as a proxy to the fraction of early-type galaxies. As discussed in Section 3.3, the results from Moffett et al. (2016a), and thus Moffett et al. (2016b), overestimate the fraction of early-type galaxies compared to the SDSS morphological catalogues. The above could be due to the inclusion of Sa galaxies into the group of early types. We used the above to argue in favour of our derived fraction of early-type galaxies based on the automated morphological classification from Huertas-Company et al. (2011).

Following Paper I, we use the Bayesian approach described previously through a MCMC method applied jointly to all the data (the CPDFs in different $M_*$ bins) to find the best-fit parameters

of the proposed functions. These are listed in Table 3. Figure 3 shows our best-fitting models for late-type galaxies compared to the CPDFs from Paper I. Figure 4 shows the same but for early-type galaxies. We notice that our best-fit parameters are very similar to those determined in Paper I.

### 4.2. The HI- *and* H₂-*to-stellar mass correlations*

Next, we explore the resulting first and second moments from our best-fitting models to the observed HI- and $H_2$-CPDFs, shown in Figure 5. The left panels of the figure present the logarithmic mean $\langle \log \mathcal{R}_j \rangle$ and its corresponding standard deviation, $\sigma_{\log \mathcal{R}_j}$, $j = $ HI or $H_2$, as a function of $M_*$ for early- and late-type galaxies as well as for all the galaxies. At low masses, the correlation of all galaxies approaches the one of the late-type galaxies, while at high-mass end it approaches early types. The above trends are just the consequence of the observed fraction of early and late types. Figure 5 shows that early- and late-type galaxies follow different $\langle \log \mathcal{R}_j \rangle - M_*$ correlations. Therefore, due to the strong bimodality of these correlations, conclusions based on some subset of galaxies as representative of all galaxies would lead to incorrect results.

In the literature, sometimes the gas-to-stellar mass relations are reported using the arithmetic mean (though the results are plotted in logarithmic diagrams). The right panel of Figure 5 shows $\log \langle \mathcal{R}_j \rangle$ versus $M_*$ from our empirical CPDFs. As is clearly seen, there are notable differences when computing different ways of averaging the distributions: *(i)* $\log \langle \mathcal{R}_j(M_*) \rangle > \langle \log \mathcal{R}_j(M_*) \rangle$, being larger the difference for the early-type galaxies;[n] and *(ii)* the standard deviations from the arithmetic mean is smaller than from the logarithmic mean.

In the left upper panel of Figure 4, we reproduce the results from (Maddox et al. 2015) for the ALFALFA galaxies with SDSS spectral and stellar mass counterparts. It is clear that the ALFALFA survey is biased towards galaxies with high HI-to-$M_*$ ratios. In

---

[n]In the case of the arithmetic mean, the contribution of low values, even if they dominate in number, could be in some cases significantly lower than higher values. Then, for the arithmetic mean the contribution of low $\mathcal{R}_j$ values is minimised contrary to the logarithmic mean of $\mathcal{R}_j$.
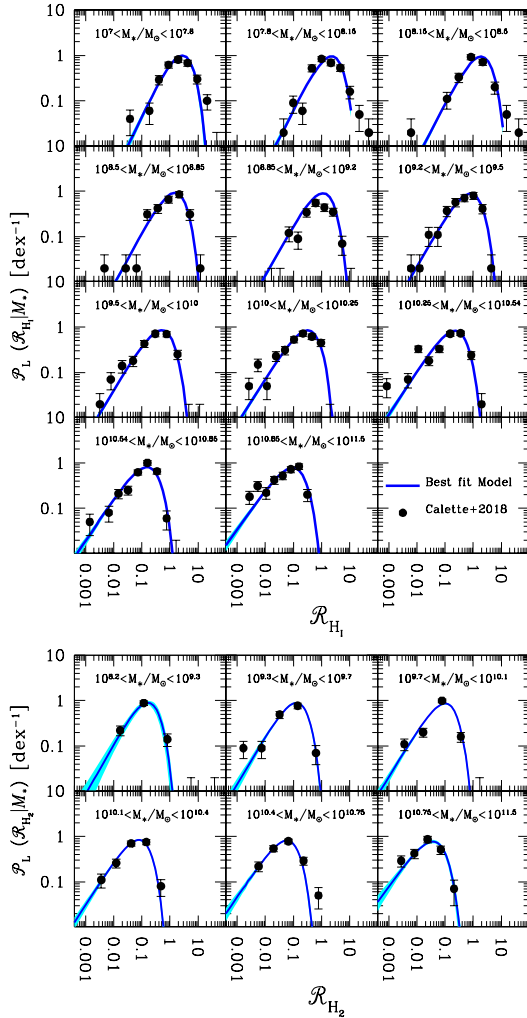
**Figure 3.** HI and $H_2$ mass CPDFs for late-type galaxies. The results for the compilation sample from Paper I are shown as filled circles with error bars. Note that the above results include non-detections since the authors used the Kaplan & Meier (1958) estimator for uncensored data in their analysis. Our best-fitting models are shown as the solid lines.

other words, the ALFALFA survey mainly detects galaxies in the upper envelope of the full distribution of $\mathcal{R}_{HI}$ (see also Maddox et al. 2015) and is dominated mostly by late types (see also e.g., Haynes et al. 2011).
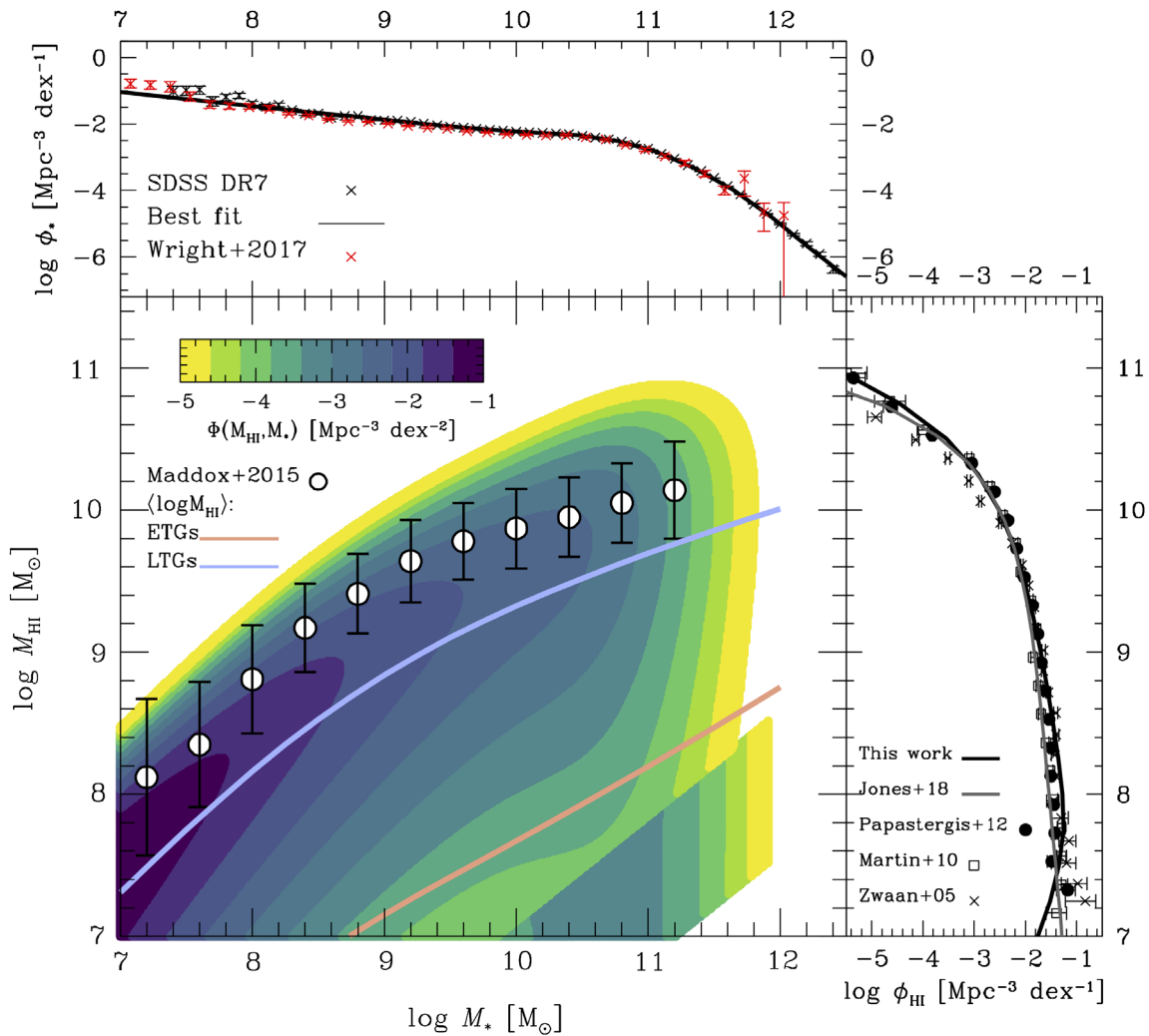
### 4.3. The bivariate mass distribution functions

Figure 6 shows the resulting bivariate stellar-HI mass distribution function, $\Phi(M_{HI}, M_*)$, see Equation (2). The colour code shows various number density levels for $\Phi(M_{HI}, M_*)$. Notice that $\Phi(M_{HI}, M_*)$ is for all galaxies, that is, it includes the contribution from early- and late-type galaxies. The discontinuity in the number density isocontours at the bottom right of the diagrams is related to contributions from the non-detections from early types. Recall that for the CPDFs of early types, we assumed an uniform function (or top-hat) for the lowest values of the gas-to-stellar mass ratios $\mathcal{R}_{HI}$, where the non-detection
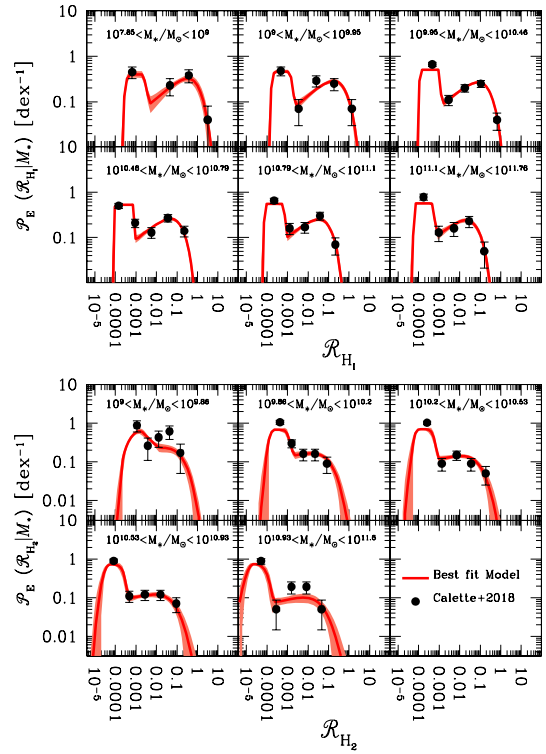


**Figure 4.** Same as Figure 3 but for early-type galaxies. Note that the CPDFs of early-type galaxies reported in Paper I account for upper limits corrected by distance selection effects when necessary and the treated with the Kaplan & Meier (1958) estimator, see Section 2.2.1.
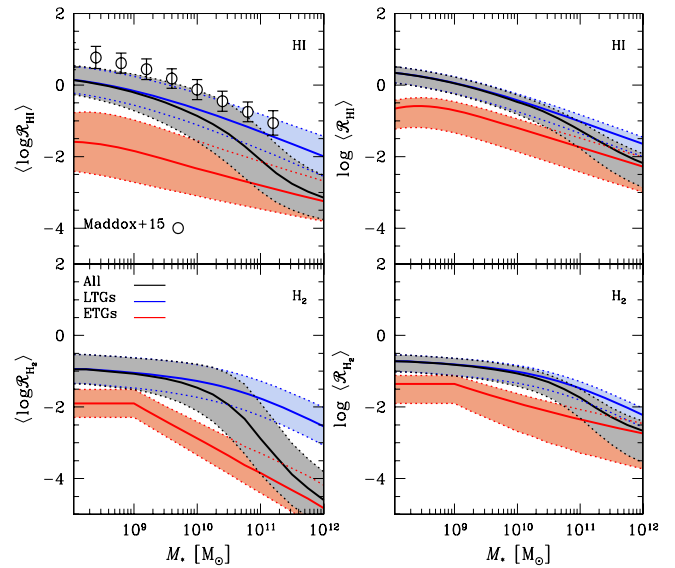


**Figure 5.** Logarithmic, left panels, and arithmetic, right panels, averaged mass ratios $\mathcal{R}_j$ as a function of $M_*$ from our analysis, with $j =$ HI, $H_2$. Blue and red lines are for early- and late-type galaxies, respectively, while the black lines correspond to all galaxies. The shaded areas show the respective standard deviations. Notice that $\log\langle\mathcal{R}_j(M_*)\rangle \geq \langle\log\mathcal{R}_j(M_*)\rangle$ and the dispersion reduces for the arithmetic mean. The open circles with error bars in the upper left panel correspond to the data from ALFALFA galaxies with SDSS spectral and stellar mass counterparts according to Maddox et al. (2015).
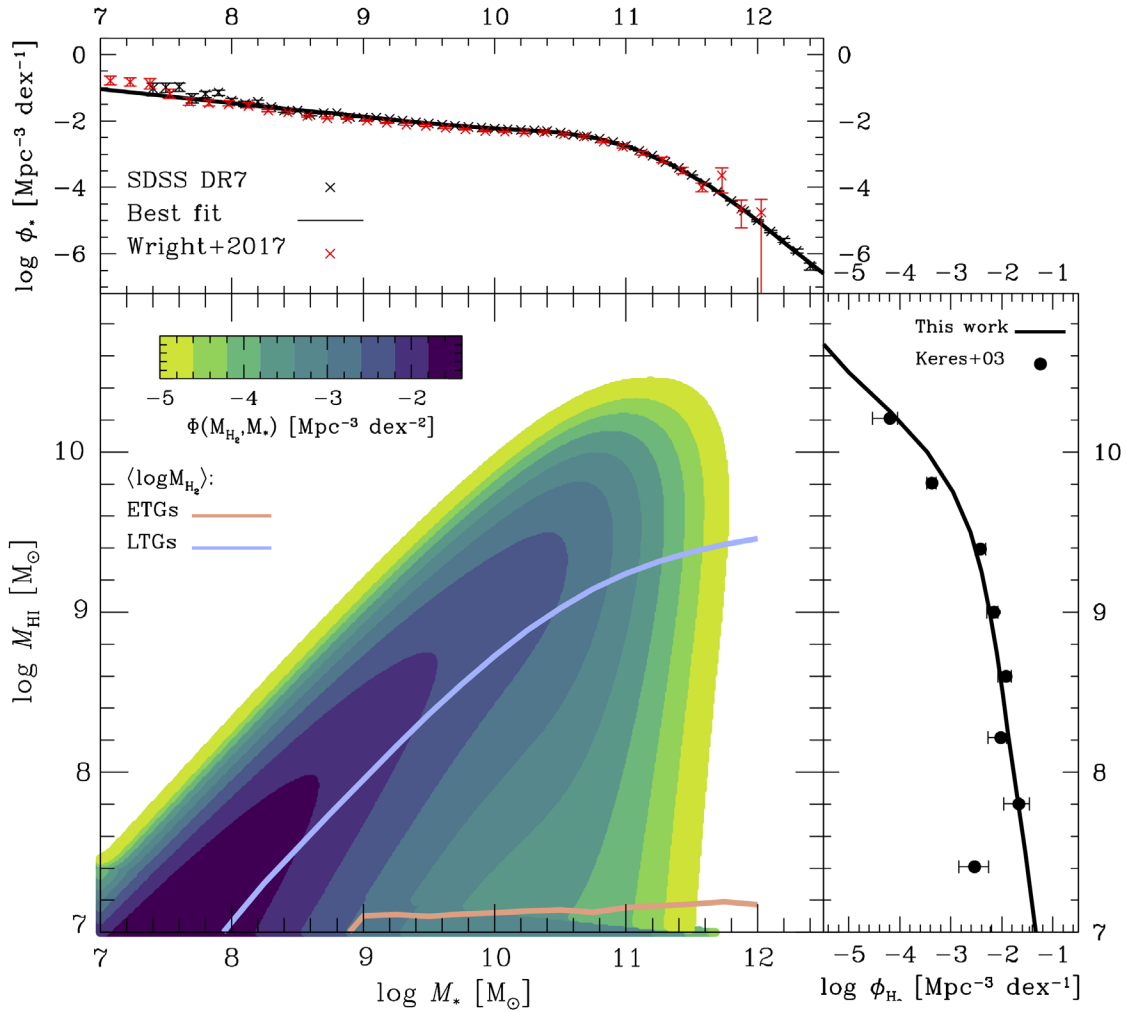
**Figure 6.** Atomic gas–stellar mass bivariate (joint) distribution function. The colour code shows various number density levels as indicated by the legends. Due to the rising slope of the MFs at low masses, most of the galaxies are located at small HI and stellar masses. Note that the discontinuity seen at the low-HI and high-stellar masses is due to the assumption of an uniform function for the lowest values of gas-to-stellar mass ratios of early-type galaxies where non-detections piled up. Recall that in our analysis, non-detections (upper limits) are included using the non-parametric estimator Kaplan & Meier (1958) for censored data in Paper I. The solid lines show the mean $\langle \log M_{HI} \rangle$ as a function of $M_*$, both for early- and late-type galaxies. The upper panel shows the GSMF which is the result of integrating the bivariate distribution function along the $M_{HI}$ axis, while the bottom right panel shows the same but for the HI MF which results from integrating along the $M_*$ axis. We compare to some previous observational determinations of the MFs and the relationship between $M_{HI}$ and $M_*$ derived in Maddox et al. (2015) for the ALFALFA survey with SDSS spectral and stellar mass counterparts.

piled up,[o] see Section 2.2.2 and Figure 4. In the bottom right and upper panels of the same figure, we present respectively our measurements of the HI MF and GSMF with the solid black lines. We compare the HI MF with *blind* HI galaxy surveys based on ALFALFA (Jones et al. 2018; Papastergis et al. 2012; Martin et al. 2010) and on HI Parkes All Sky Survey (HIPASS) (Zwaan et al. 2005). While in the next subsection, we discuss in more detail the comparison with previous works, for the moment we note that our total HI MF is in good agreement with the above direct observations. In the case of ALFALFA, this is a revealing result given the strong selection bias of this survey towards HI-rich and late-type galaxies as seen in Figure 6 (open circles reproduce the results from

Maddox et al. 2015, see also the discussion of the previous subsection and Figure 4). As we will discuss in the next section, the above reflects that the total HI MF is dominated by late-type galaxies.

Figure 6 explicitly shows the contribution of galaxies of different stellar masses to the HI MF. Particularly, we observe that the low-mass end of the HI MF is composed mainly by low $M_*$ galaxies but there is also a non-negligible contribution from a population of high $M_*$ galaxies. Most of these high $M_*$ galaxies are early-type (quenched) galaxies for which there is a significant fraction of non-detections ($\sim$55%). In Paper I, we included non-detections for the determination of the HI-CPDF based on methods of censored data (Kaplan & Meier 1958). Nonetheless, the contribution of non-detections is only marginal because the fraction of early-type galaxies at those masses is low, see the bottom panel of Figure 2. In addition, Figure 6 shows that the completeness limit in the HI MF, due to our stellar mass limit of $M_* = 10^7 M_\odot$, is at

---

[o]Note that the top-hat is not the result of applying the Kaplan & Meier (1958) estimator as we a posteriori redistributed the lowest values of $\mathcal{R}_{HI}$ (including upper limits) into a uniform function.

**Figure 7.** Same as Figure 6 but for molecular gas. Note that while there are more non-detections for $H_2$ observations, these are mostly from early-type galaxies that represent only a small fraction overall in the $H_2$ mass bivariate distribution function. We also compared to previous determinations from Keres et al. (2003).

$M_{HI} \sim 10^8 M_\odot$ (see below), which excludes a large region of galaxies with non-detections.

Similarly to Figure 6, Figure 7 presents the resulting bivariate stellar–$H_2$ mass distribution function for all galaxies and the mean $\langle \log M_{H2} \rangle$ for early- and late-type galaxies. The resulting total $H_2$ MF is shown with the solid line in the bottom right panel and compared to the Keres et al. (2003) $H_2$ MF based on the CO luminosity function. At the low-mass end, there is a substantial population of galaxies with non-detections, roughly $\sim$78%, which are mostly early-type galaxies. As above, non-detections have been included in the statistical analysis of the H2-CPDFs, for details see Paper I. Nevertheless, from the contour density level, their contribution is marginal. Finally, we can conclude that our $H_2$ MF is complete for $M_{H2} \gtrsim 10^7 M_\odot$.

### 4.4. The mass functions

Next, we discuss in detail the MFs presented above. In particular, we focus on the determinations separately for early- and late-type galaxies based on the morphology classification described in Section 3.3. The various panels in Figure 8 present the MFs for atomic and molecular gas, total cold gas, and baryons, as indicated by the labels. In all the panels, the MFs for late-type galaxies are shown as blue filled circles with error bars, while for early-type galaxies are shown as red circles with error bars. These symbols are for a stellar mass limit of $M_* > 10^7 M_\odot$. Instead, we use blue/red open circles when the MFs is incomplete. We also calculate the MFs in the hypothetical case of a mass limit of $M_* = 0$ shown with the blue/red dashed lines. Note that the total MFs for HI and cold gas are not plotted. This is because in these cases the MFs of late-type galaxies are hardly distinguishable from the total one at all masses. We also omit the total $H_2$ MF. This is because it is hardly distinguishable from the MF of late-type galaxies at $M_{H2} \lesssim 2 \times 10^{10} M_\odot$ while for larger masses it is indistinguishable from the MF of the early-type galaxies. In the case of the baryon masses, the total MF is plotted with black filled circles.

We note that our determinations for the MFs are the result of the convolution between random errors and the intrinsic MFs, similarly as it happens with the direct observational determinations of MFs. In Section 5.1, we discuss the impact from random errors and present the intrinsic MFs, after deconvolving by these errors.
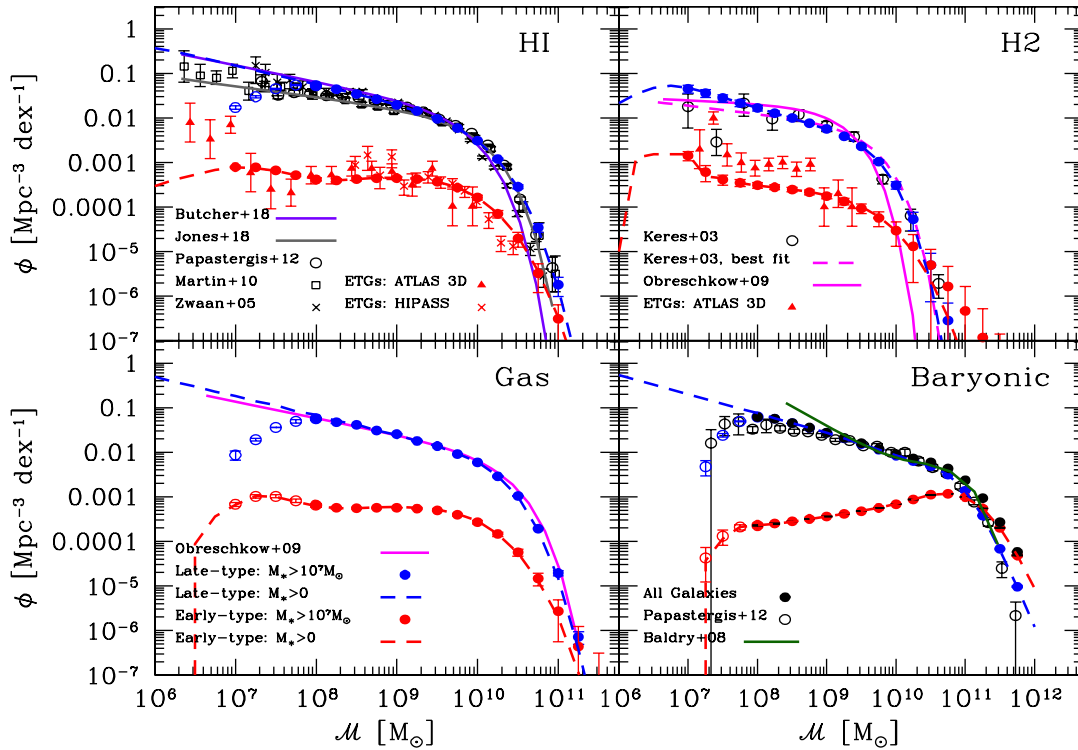
**Figure 8.** Results on the galaxy MFs of early- and late-type galaxies for atomic gas, left upper panel, molecular gas, right upper panel, cold gas, bottom left, and baryons, bottom right panel. In all the panels, late-type galaxies are shown as the blue circles with error bars, while early-type galaxies are shown as the red circles with error bars, when using a stellar mass limit of $M_* = 10^7 M_\odot$. Filled blue/red circles indicate when the MFs are complete, while open circles clearly show that the MFs became incomplete. The dashed lines are for MFs when using a stellar mass limit of $M_* = 0$. The total MFs for HI and cold gas are not shown because they are practically indistinguishable from the respective MFs of late-type galaxies. Our results are in good agreement with observational determinations of the total MFs. For only early-type galaxies, we compare our results with those from the ATLAS 3D sample (red triangles). While we observe some tension, we suspect that selection effects are more likely to artificially increase the amplitude of their MF at low masses.

### 4.4.1. The HI MF

We compare our results with previous direct observational determinations of the *total* HI MFs. According to our results, late-type galaxies dominate the HI MF, even at the highest masses, so that it is practically equivalent to the total HI MF. In Figure 8, we reproduce the best fit to the HI MF from Jones et al. (2018) based on the final catalogue of the blind HI ALFALFA Survey, dark grey solid line. The violet solid line shows the best-fit model reported in Butcher et al. (2018) to the HI MF from the Nançay Interstellar Baryons Legacy Extragalactic Survey, which is a project that complemented recent and/or ongoing large blind HI surveys. Results from Papastergis et al. (2012) and Martin et al. (2010) based on the 40% sample of the ALFALFA Survey are shown respectively with black open circles and squares. The skeletal symbols with error bars show the results from Zwaan et al. (2005), who used the HIPASS. Note that our HI MF for late-type galaxies, which dominates the total MF, is in good agreement with direct inferences from HI blind surveys, particularly those based on ALFALFA as discussed in Figure 6. As for the HI MF of early-type galaxies, we plot the determinations from the ATLAS 3D (Serra et al. 2012) and HIPASS (Lagos et al. 2014) samples shown as the red solid triangles and skeletal symbols, respectively.

Our resulting HI MFs are in good agreement with direct determinations from radio observations. This is particularly true for HI masses above the completeness limit corresponding to our $M_*$ limit for the GSMF. These masses are $M_{HI} \sim 10^8 M_\odot$ for

late-type galaxies and $M_{HI} \sim 10^7 M_\odot$ for early-type galaxies. Even when extrapolating to a limit mass of $M_* = 0$, at the low-mass end, we find a good agreement with direct determinations, though the early-type galaxies from the ATLAS 3D sample present a higher amplitude for masses below $10^7 M_\odot$. However, those extrapolations should be taken with caution as it is not clear whether we can extrapolate our bivariate distribution functions to such low masses. In conclusion, we consider that the remarkable consistency between *our HI MFs and radio blind surveys above* $M_{HI} \sim 10^8 M_\odot$ *is reassuring* and validates the HI-CPDFs determined in Paper I. Recall that the observational data used in that paper were derived from various heterogeneous samples, affected by many selection effects, including those related to the non-radio detections. Therefore, the agreement between the HI MF with that of the blind radio observations is far from trivial, unless adequate corrections are introduced and the data are adequately analysed from the statistical point of view.

### 4.4.2. The $H_2$ MF

In the upper right panel of Figure 8, we present the results for the $H_2$ MF. The $H_2$ MF is largely dominated by late-type galaxies below $M_{H_2} \sim 2 \times 10^{10} M_\odot$, but for larger masses, early-type galaxies are more abundant. In the same panel, we reproduce the total $H_2$ MF from Keres et al. (2003), who used a CO luminosity function from FIR and $B-$band limited galaxy samples and adopting a constant CO-to-$H_2$ conversion factor of $\alpha_{CO} = 4.76$, open black

circles with error bars. The dashed line shows the best fit to a Schechter function derived in Obreschkow & Rawlings (2009). Additionally, we show the results from the ATLAS 3D sample for early-type galaxies (Lagos et al. 2014) with the filled triangles. The magenta solid line shows the results from Obreschkow & Rawlings (2009), who derived the $H_2$ MF by introducing a phenomenological model for the $H_2$-to-HI mass ratio that depends on the galaxy morphological type and its total cold gas mass.

When comparing to the $H_2$ MF from Keres et al. (2003), we observe a good agreement with our results. At the low-mass end, though, the Keres et al. (2003) MF seems to be slightly shallower than ours. It is not clear the origin for this discrepancy. One possibility is due the constant $\alpha_{CO}$ factor employed by the authors. Based on previous empirical studies, Paper I showed that ignoring the dependence of $\alpha_{CO}$ with $M_*$[P] flattens the low-mass end of the $H_2$ MF, consistent with the results from Keres et al. (2003). Another possibility is an effect of the incompleteness of the CO luminosity function. As for Obreschkow & Rawlings (2009), our results are consistent for masses below $M_{H_2} \sim 3 \times 10^9 M_\odot$. For larger masses, the Obreschkow & Rawlings (2009) MF falls much stepper than ours. Similar to our analysis, Obreschkow & Rawlings (2009) used the relationships between galaxy properties to derive their MF. As mentioned above, their phenomenological model employed the dependence of the $H_2$-to-HI mass ratio with morphology and cold gas mass. While the above differences could be arguably referred to the use of different estimators for the $H_2$ gas masses, it could be also an indication that random errors are larger when using only one galaxy parameter. Recall that in this paper, we are using $M_*$. In Section 5.1, we will show that after deconvolving from random errors, our intrinsic $H_2$ MF becomes steeper at the high-mass end and more consistent with the derivation from Obreschkow & Rawlings (2009). Nonetheless, it is difficult to conclude the origin of the above differences given the different nature of the models employed in both studies.

As for early-type galaxies, our results are consistent with those from the ATLAS 3D (Lagos et al. 2014) at the high-mass end but they lie slightly below at the low-mass end. It is unclear the reason of the above discrepancy for low-$M_{H_2}$ early-type galaxies, though large-scale and environmental selection effects could boost the inferences of the MF when using the $1/V_{max}$ estimator, see for example, Appendix C and Baldry et al. (2008). Recall that in the case of HI, the ATLAS 3D also presents an slightly enhancement at the low-mass end of the HI MF. Thus, selection effects are more likely to artificially increase the amplitude of the MF for low-mass galaxies in the ATLAS 3D sample.

### 4.4.3. The gas and baryonic MFs

Similarly to the HI MF, the cold gas MF is completely dominated by late-type galaxies, even at the high-mass end. In Figure 8, we compare our results with the phenomenological determination from Obreschkow & Rawlings (2009, pink solid line). These authors combined their inference of the $H_2$ MF with the HI MF from Zwaan et al. (2005) to derive the gas MF. Despite the differences mentioned above for the Obreschkow & Rawlings (2009) $H_2$

MF, their total cold gas MF is in excellent agreement with our one. This is not surprising as it is just reflecting that HI is much more abundant than $H_2$.

Finally, we show our results for the baryonic MFs in the bottom right panel of Figure 8. The baryonic MF is very similar to the GSMF at the high-mass end but at the low-mass end, it is steeper as the contribution of cold gas becomes more relevant. On the other hand, late-type galaxies dominate the baryonic MF for $M_{bar} < 10^{11}$ $M_\odot$, while at the high-mass end, early-type galaxies are more abundant than the late-type ones. We reproduce with the green solid line the baryonic MF from Baldry et al. (2008). These authors combined the GSMF from the low-z survey of the SDSS DR4, the same galaxy survey used here for low masses, with a closed-box model and the mass–metallicity relation to derive cold gas masses for their baryonic MF. The open black circles show the MF from Papastergis et al. (2012), who defined baryonic mass as $M_{bar} = 1.4 \times M_{HI} + M_*$. We notice that these previous baryonic MF determinations are in good agreement with our results at the mass range $\sim 2 \times 10^9 - 2 \times 10^{11}$ $M_\odot$, while for lower and higher masses there are some differences, which are easy to understand.

The MF from Baldry et al. (2008) is steeper than our MF at low masses. This might be a consequence of the fact that the Baldry et al. (2008) GSMF itself is steeper compared to other determinations, for example, Baldry et al. (2012). As these authors discuss, the disagreement between the Baldry et al. (2008) and Baldry et al. (2012) GSMFs is just the result of different flow models for distances, which affect significantly to nearby low-mass galaxies. Recall that our GSMF has been corrected to be consistent with the flow model of Tonry et al. (2000). Additionally, the gas masses in Baldry et al. (2008) were actually obtained from a close-box model constrained with the empirical mass–metallicity relation. The combination of these two assumptions are likely the result of a steep baryonic MF at low masses, which differs from our results and those of Papastergis et al. (2012).

Regarding the high-mass end, our baryonic MF falls shallower than those of Papastergis et al. (2012) and Baldry et al. (2008). This is because, our GSMF is shallower. As discussed in Section 3, there are two main systematic effects that could lead to different GSMFs, mass-to-light ratios, and the determination of galaxy surface brightness (especially due to sky subtraction problems). Both effects are likely to affect the high-mass end of the baryonic MF. In addition, due to the small volumes of the surveys used in Baldry et al. (2008) and Papastergis et al. (2012), cosmic variance enhances the differences.

### 4.5. Cosmic density parameters and relevant timescales

#### 4.5.1. Cosmic density parameters

The cosmic density parameter measures the average mass density of some matter species with respect to the critical density, $\rho_c$. Here, we determine the mass density in stars, HI, $H_2$, cold gas, and baryons that are locked inside galaxies using their MFs. The differential mass density function $d\rho_j(M_j)$ for some galaxy mass component $M_j$ in the mass range $\log M_j \pm d \log M_j/2$ is $d\rho_j(M_j) = M_j \times \phi_j(M_j) d \log M_j$, where $\phi_j$ is in units of Mpc-3 dex-1. Thus, the cosmic mass density is given by:

$$\rho_j = \int_{-\infty}^{\infty} d\rho_j(M_j), \tag{26}$$

---

[P]In Appendix C of Paper I, we have constrained the CO-to-$H_2$ conversion factor to be mass-dependent: $\log (\alpha_{CO}) = 0.15 + 0.35[1 + 0.1(3 \times 10^{10}/M_*)^{0.64}]$ down to $M_* = 10^8$ $M_\odot$ and for lower masses the value of $\alpha_{CO}$ remains constant. Therefore, $\alpha_{CO}$ increases as $M_*$ decreases saturating to a value of $\approx 250$ for $M_* < 10^8$ $M_\odot$. This is due to the empirical dependences of $\alpha_{CO}$ on the gas-phase metallicity, and the dependence of the latter with $M_*$

**Table 4.** Cosmic density of HI, $H_2$, gas, stars, and baryons for all, LTGs and ETGs. The fraction of each component is denoted as $f_j = \Omega_j / \Omega_{bar,U}$ with $\Omega_{bar,U} = 0.048$.

|  | $\Omega_{H_2}/10^{-4}$ | $f_{H_2}$ | $\Omega_{HI}/10^{-4}$ | $f_{HI}$ | $\Omega_{gas}/10^{-4}$ | $f_{gas}$ | $\Omega_*/10^{-4}$ | $f_*$ | $\Omega_{bar}/10^{-4}$ | $f_{bar}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| All | $0.86 \pm 0.05$ | 0.18% | $4.24 \pm 0.10$ | 0.88% | $6.85 \pm 0.92$ | 1.43% | $20.40 \pm 0.08$ | 4.25% | $26.01 \pm 0.13$ | 5.42% |
| LTG | $0.82 \pm 0.04$ | 0.17% | $4.09 \pm 0.10$ | 0.85% | $6.59 \pm 0.89$ | 1.37% | $13.20 \pm 0.05$ | 2.75% | $18.25 \pm 0.12$ | 3.80% |
| ETG | $0.04 \pm 0.01$ | $\sim$0.01% | $0.15 \pm 0.02$ | 0.03% | $0.21 \pm 0.03$ | 0.04% | $7.21 \pm 0.03$ | 1.50% | $7.76 \pm 0.37$ | 1.62% |

LTG = late-type galaxies; ETG = Early-type galaxies
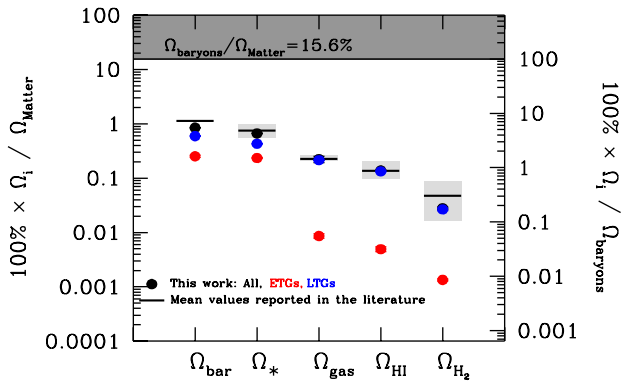


**Figure 9.** Density parameter $\Omega$ of HI, $H_2$, cold gas, and baryonic mass locked in all galaxies as well as in early- and late-type galaxies (coloured filled circles; the errors are smaller than the circle size). The $\Omega$ parameter values are reported as fractions in per cent of the universal matter (left axis) and baryonic (right axis) densities. The grey boxes show the range of values from previous determinations and the horizontal lines correspond to the mean of these values.

with the cosmic density parameter

$$\Omega_j = \frac{\rho_j}{\rho_c}, \tag{27}$$

where the critical density is $\rho_c = 2.775 \times 10^{11} h^{-1} M_\odot / (h^{-1} \text{Mpc})^3 = 1.2756 \times 10^{11} M_\odot / \text{Mpc}^3 h_{67.8}^2$.[q] The limits of integration in Equation (26) reflect that we are considering all the spectrum of masses from galaxies. In reality, this is not possible, due to completeness limits in galaxy samples. Here, we consider the following mass limits for all our components: $M_{inf} = 10^7$ $M_\odot$ and $M_{upper} = 10^{12.6}$ $M_\odot$. We notice that using values smaller than $M_{inf}$ and/or larger than $M_{upper}$ do not substantially change our results. This is because the multiplicity functions, $\propto M_j \times \phi_j(M_j)$, have a maximum around the knee of the MFs.

Figure 9 shows the different values of each $\Omega_j$ corresponding to all galaxies and separately for early- and late-type galaxies, listed in Table 4. The $\Omega_j$ values are presented as the fractions in per cent with respect to the universal matter density ($\Omega_m = 0.307$, left axis) and the universal baryonic density ($\Omega_{bar,U} = 0.048$, right axis). To estimate errors in our cosmic density parameters, we use all our MCMC models for the HI-CPDF and H2-CPDF in addition of all our MCMC models to the fit of the GSMFs. We found that the largest uncertainties arise from the uncertainties in the CPDFs.

In the past, there have been some assessments of the cosmic density parameters at $z \sim 0$. Usually, these studies do not report cosmic density parameters for different populations and for different components at the same time. As mentioned in the Introduction, it is important to distinguish between different populations given that late- and early-type galaxies follow different

formation histories. Studies close to ours are the ones by Fukugita, Hogan, & Peebles (1998), Fukugita & Peebles (2004), and Read & Trentham (2005). Below, we present and compare our results with many previous determinations from the literature.

- HI *cosmic density:* The atomic hydrogen in late-type galaxies is $\sim$27 times larger than in early-type galaxies, which means that $\sim$96% of HI mass is in late-type galaxies. Previously, Zwaan et al. (2005), Read & Trentham (2005), Martin et al. (2010), Braun (2012), Delhaize et al. (2013), Hoppmann et al. (2015), Butcher et al. (2018), and Jones et al. (2018) have derived the HI cosmic density parameter using either blind HI galaxy surveys (HIPASS and ALFALFA) or indirect techniques. The mean value from these determinations is $\Omega_{HI} = 4.2 \times 10^{-4}$ with a lower bound of $\Omega_{HI}^- = 3 \times 10^{-4}$ and an upper bound of $\Omega_{HI}^+ = 6.2 \times 10^{-4}$, shown as the grey box in Figure 9[r]. Recently, using a spectral stacking technique and from Westerbrook Synthesis Radio Telescope observations of 1895 galaxies crossed with the SDSS, Hu et al. (2019) found $\Omega_{HI} = (4.15 \pm 0.26) \times 10^{-4}$. Our determined value is in good agreement with these previous determinations, in particular with the latter one.

- $H_2$ *cosmic density:* The molecular hydrogen cosmic density in late-type galaxies is $\sim$21 times larger than in early-type galaxies. This implies that $\sim$95% of $H_2$ mass is in late-type galaxies. Using the CO surveys from Young et al. (1995), Keres et al. (2003) determined that $\Omega_{H2} = (1.64 \pm 0.63) \times 10^{-4}$, while from the observations in CO from Maeda, Ohta, & Seko (2017) they report $\Omega_{H2} = 0.51 \times 10^{-4}$. Obreschkow & Rawlings (2009) used a phenomenological model to derive $\Omega_{H2} = (1.01 \pm 0.39) \times 10^{-4}$. Read & Trentham (2005) find that $\Omega_{H2} = 2.68 \times 10^{-4}$. The above ranges of values are shown with grey box in Figure 9. As can be seen, our results are consistent with the range of determinations described above, especially with the results from Obreschkow & Rawlings (2009).

- *Cold gas cosmic density:* Most of the cold gas is located in late-type galaxies, $\sim$96%. Keres et al. (2003) found that $\Omega_{gas} = (6.34 \pm 1.62) \times 10^{-4}$, which includes the resulting abundance of HI from Zwaan et al. (1997). Obreschkow & Rawlings (2009) used their best phenomenological model to the $H_2$-to-HI ratio with the HIPASS results from the Zwaan et al. (2005) sample to derive $\Omega_{gas} = (6.49 \pm 1.18) \times 10^{-4}$, while using the values for HI and $H_2$ masses from Read & Trentham (2005), we calculate that $\Omega_{gas} = 7.95 \times 10^{-4}$ after correcting from helium and heavier metals. Our value of $\Omega_{gas} = (6.85 \pm 0.92) \times 10^{-4}$ is consistent with the above results.

- *Stellar cosmic density:* The stellar cosmic density in late-type galaxies that we derive from the SDSS is approximately $\sim$1.8 larger than in early-type galaxies. Thus, $\sim$64% of the mass in stars at $z \sim 0.1$ is in late-type galaxies. From the compilation by Madau & Dickinson (2014), the stellar cosmic density lies

---

[q] We use this symbol to emphasise that $H_0 = 67.8$ km s-1 Mpc-1 in our cosmology.

[r] All the values for the papers listed above have been renormalised to a units of $h_{67.8}^{-1}$.

within $\Omega_* = (28.06 - 17.71) \times 10^{-4}$, while the derivations from Wright et al. (2017) and Baldry et al. (2008) are respectively $\Omega_* = 17.14 \times 10^{-4}$ and $\Omega_* = 29.73 \times 10^{-4}$. Our result for the cosmic density for all galaxies, $\Omega_* = (20.20 \pm 0.08) \times 10^{-4}$, is consistent with the above values.

- *Baryonic cosmic density:* Finally, we find that there is a factor of $\sim 2.4$ more baryons in late-type galaxies than in early types, and thus, $\sim 71\%$ of the baryons are in late-type galaxies. Read & Trentham (2005) found that $\Omega_{bar} = 35 \times 10^{-4}$ which is a factor of $\sim 1.3$ larger than our results. We find a cosmic density parameters ratio of $\Omega_* / \Omega_{bar} \approx 1.3$. Finally, our baryon density parameter is $\approx 5.4\%$ of the universal baryon fraction, $f_{bar,U} = 0.156$, or equivalently $\sim 18$ times lower than $f_{bar,U}$. Most of the baryons are definitively not locked inside galaxies.

### 4.5.2. Cosmic timescales

We are now in a position to derive some relevant cosmic timescales, such as the mean galaxy depletion times. We focus on late-type galaxies because most of the star formation occurs in those galaxies. To do so, we use the observed cosmic star formation rate (CSFR) at $z \sim 0.1$. According to Madau & Dickinson (2014), who derived the CSFR from far-UV and IR rest frame luminosities, the CSFR is $\dot{\rho}_* \sim 90 \times 10^{-4}$ M$_\odot$ yr$^{-1}$ Mpc$^{-3}$ after correcting to a Chabrier (2003) IMF. Unfortunately, the authors report the CSFR for all galaxies but not divided into different groups. The recent study by Sánchez et al. (2019), based on the fossil record analysis of a sample of more than $4 \times 10^4$ galaxies from the SDSS MaNGA survey, report similar values to the the total CSFR of $\dot{\rho}_* = 114.82 \pm 67.61 \times 10^{-4}$ M$_\odot$ yr$^{-1}$ Mpc$^{-3}$ or $\dot{\Omega}_* = \dot{\rho}_*/\rho_{crit} = (9 \pm 5) \times 10^{-14}$ yr$^{-1}$ corrected to a Chabrier (2003) IMF. The authors also derived the CSFRs for star-forming galaxies; $\dot{\Omega}_{*,SFG} = (6.5 \pm 3.8) \times 10^{-14}$ yr$^{-1}$. In the following, we use their value for star-forming galaxies as a representative determination for late-type morphologies, that is, $\dot{\Omega}_{*,L} \approx \dot{\Omega}_{*,SFG}$.

We begin our discussion by estimating the mean molecular hydrogen depletion time of late-type galaxies, $\bar{t}_{dep,L}(H_2) = \Omega_{H_2,L}/\dot{\Omega}_{*,L}$. The $H_2$ depletion timescale is defined as the time at which a galaxy (or a molecular cloud) would consume its $H_2$ gas reservoir by forming stars at the current SFR. From our cosmic density parameters, we find that $\bar{t}_{dep,L}(H_2) \approx 1.3$ Gyrs. This is consistent with the mean depletion time $\bar{t}_{dep}(H_2) = 0.96$ Gyr reported in Saintonge et al. (2017) for star-forming galaxies in a volume complete sample. Note, however, that for local individual galaxies the molecular depletion time could vary from $\sim 0.9$ to 3 Gyrs (e.g., Kennicutt 1998; Bigiel et al. 2008; Leroy et al. 2008; 2013). Also, we estimate the mean total cold gas depletion time of late-type galaxies, $\bar{t}_{dep,L}(gas) = \Omega_{gas,L}/\dot{\Omega}_{*,L}$, and find $\bar{t}_{dep,L}(gas) \approx 10.14$ Gyrs, that is, $\sim 8$ times larger than for the molecular gas component. The values we find for these two timescales are consistent with the proposal that, on average, for local late-type galaxies, *(i)* the global conversion of molecular gas into stars is inefficient (recall that the $H_2$ depletion times of observed local star-forming regions are actually 40–500 Myr, for example, Lada, Lombardi, & Alves 2010; Lada et al. 2012); and *(ii)* the global conversion of atomic to molecular hydrogen gas is also inefficient, or equivalently, the molecular cloud formation efficiency is low. Thus, the mean star formation efficiency, SFE, of local late-type/star-forming galaxies is low despite the fact that they contain a considerable amount of interstellar gas; according to Table 4, on average approximately 36% of the baryons in these galaxies are in form of cold gas.

According to Leroy et al. (2008), the SFE of a galaxy is the inverse of the neutral H gas depletion time, that is, the time required for current star formation to consume the neutral H reservoir. The SFE can be estimated as the product of the net efficiency of converting $H_2$ gas into stars, $SFE(H_2) = 1/t_{dep}(H_2)$, and the net efficiency of molecular cloud formation given by the mass fraction of $H_2$ with respect to the total neutral H mass in the galaxy, that is, $M_{H_2}/(M_{HI} + M_{H_2})$. Thus, using our estimations of the cosmic parameters for late-type galaxies, we calculate the cosmic (mean) SFE of late-type galaxies as:

$$SFE_L(H) = SFE_L(H_2) \times \frac{\Omega_{H_2,L}}{\Omega_{HI,L} + \Omega_{H_2,L}}$$

$$= \frac{1}{\bar{t}_{dep,L}(H_2)} \times \frac{\Omega_{H_2,L}}{\Omega_{gas,L}/1.4} = 1.4 \times \frac{\dot{\Omega}_{*,SF}}{\Omega_{gas,L}}$$

$$= \frac{1.4}{\bar{t}_{dep,L}(gas)} = 1.38 \times 10^{-10} yr^{-1}. \quad (28)$$

The inverse of this efficiency is the cosmic neutral H depletion time, $\bar{t}_{dep,L}(H) \approx 7.25$ Gyrs. Note that the relationship between the SFE based on the neutral H gas reservoir and the SFE based on the total cold gas reservoir is $SFE_L(H) = 1.4 \times SFE_L(gas)$, or equivalently, $\bar{t}_{dep,L}(H) = \bar{t}_{dep,L}(gas)/1.4$. The factor 1.4 takes into account He and metals.

We calculate also the cosmic SF timescale of late-type galaxies, which is given by the inverse of the cosmic-specific SFR, $\bar{t}_{SF,L} = \Omega_{*,L}/\dot{\Omega}_{*,L} \approx 20.3$ Gyrs; this is a factor of $\sim 1.5$ larger than the present age of the Universe. The cosmic SF timescale can be understood as the time required for the current cosmic SFR density to double the current cosmic stellar mass content. Interestingly enough, the ratio $\bar{t}_{dep,L}(H)/\bar{t}_{SF,L} = (\Omega_{H_2} + \Omega_{HI})/\Omega_* = 0.36$, that is, the gas reservoir of late-type galaxies has not yet been dramatically consumed by star formation. Including He and metals in the gas reservoir, this ratio increases to $\sim 0.5$.

## 5. Discussion

In this paper, we employed a statistical approach that allows to project the observed HI- and $H_2$-CPDFs into their corresponding MFs, when using the GSMF as an interface or pivotal function, Section 2. Additionally, the cold gas and baryon MFs are obtained from the above. Our empirical approach makes use of the following observational data as input:

1. The local GSMF over a large dynamical range and separated into early- and late-type galaxies.
2. The observed CPDFs of HI and $H_2$ as a function of $M_*$, both for early- and late-type galaxies.

As a result, our approach provides a fully self-consistent and complete empirical description of the demographics of the local population of early- and late-type galaxies for a broad mass range. Furthermore, by construction, our MFs are derived separately for early- and late-type galaxies. As discussed in Section 4.4, our HI and $H_2$ MFs are actually consistent with several previous determinations from radio *blind* or optically/infrared (selected) galaxy samples. Actually, the above level of agreement is not trivial due to the chain of assumptions and corrections for the data sets we used here and in Paper I, and it reinforces the robustness of the observational information employed. Note, however, that the above agreement is only valid above our completeness limit for

the GSMF of $M_* = 10^7 M_\odot$, which corresponds to a completeness limit of $M_{HI} \sim 10^8 M_\odot$ and $M_{H_2} \sim 10^7 M_\odot$, respectively, for the HI and $H_2$ MFs. In that regards, we are unable to constrain the very low-mass end of the HI and $H_2$ MFs.

Below, we highlight aspects that we consider are relevant for the success of our empirical approach:

- The HI and $H_2$ CPDFs for early- and late-type galaxies. We used the CPDFs from Paper I, where we derived the CPDFs from a compilation of many incomplete and inhomogeneous samples, carefully homogenised to a common IMF, cosmology, CO-to-luminosity conversion factor, and accounting for selection biases.
- The effect from upper limits in radio surveys. In addition to the above-mentioned homogenisation and corrections, it was important to take into account the upper limits reported in the original sources, when radio detections were not achieved. The fraction of non-detections in the compilation from Paper I was non negligible, especially for early-type galaxies. Non-detections were corrected by distance-sensitivity effects. Instead of ignoring radio non-detection or using the upper limits as the true values, as is commonly done in the literature, we derived the CPDFs by including them in our statistical analysis based on the non-parametric Kaplan–Meier estimator for censored data.

Next, we briefly discuss below some potential caveats on our approach. Over the next subsections, we will discuss them in more detail and show that they do not affect our main conclusions.

- The assumption that the HI and $H_2$ masses are two independent random variables. In reality, this is not true, for example, Obreschkow & Rawlings (2009) showed that the $H_2$-to-HI mass ratio depends on the morphological type. Note, however, that this was partially taken into account in our approach in a statistical sense. Recall that we use the observed mass CPDFs separately for early- and late-type galaxies, that is, the dependence with morphology is roughly included, as shown in Paper I, see Figure C3 from that paper and Figure 5.
- Differences on the mass-to-light ratios. Figure A1 in Appendix A shows that the different mass-to-light ratios used to estimate $M_*$ lead to different GSMFs, with differences up to $0.5 - -1$ dex in number densities at the high-mass end (see also Bernardi et al. 2017). While we choose to use the geometric mean over the five mass-to-light ratios described in Appendix A, one could naturally question that the agreement of our HI and $H_2$ MFs with the observed ones is relative because using a different GSMF could result in different MFs. In Section 5.1, we explore and quantify the impact of systematics from varying mass-to-light ratios and show that its effect is marginal in the obtained HI and $H_2$ MFs.
- Random errors from stellar mass estimates. Inevitably, random errors propagate to our MFs resulting in a Eddington (1940) bias effect. Thus, the comparison with the results based on radio surveys is not trivial as they do not suffer of an Eddington (1940) bias effect due to $M_*$ errors. Nonetheless, measurements of the HI and $H_2$ masses are also subject to random errors. In Subsection 5.1.1, we deconvolve our MFs with the random errors, not only as a method to compare with results from radio surveys but also for obtaining the intrinsic MFs to be used to constrain the predictions from galaxy formation models.

- The morphological classification from the SDSS DR7. To derive our GSMF separated into early- and late-type galaxies, we used the morphological classification based on the Huertas-Company et al. (2011) vector machine analysis of the SDSS DR7. As shown in Figure 2, we find that the obtained early- and late-type GSMFs using this classification are consistent with other determinations for the SDSS but disagree with those based on the visual classification from the GAMA survey. While we explore in detail this effect in Appendix C.4, we do not include it as one of the main sources of uncertainty.

We conclude this section by emphasising the robustness of the MFs derived when combining observational gas-to-stellar mass correlations from small data sets with the GSMF, (see also, Lemonias et al. 2013; Butcher et al. 2018). While this is an indirect method to study the demographics of the galaxy distribution, it is a valid and valuable approach that gives results that are comparable to direct observations and generalise them into a full bivariate distribution.

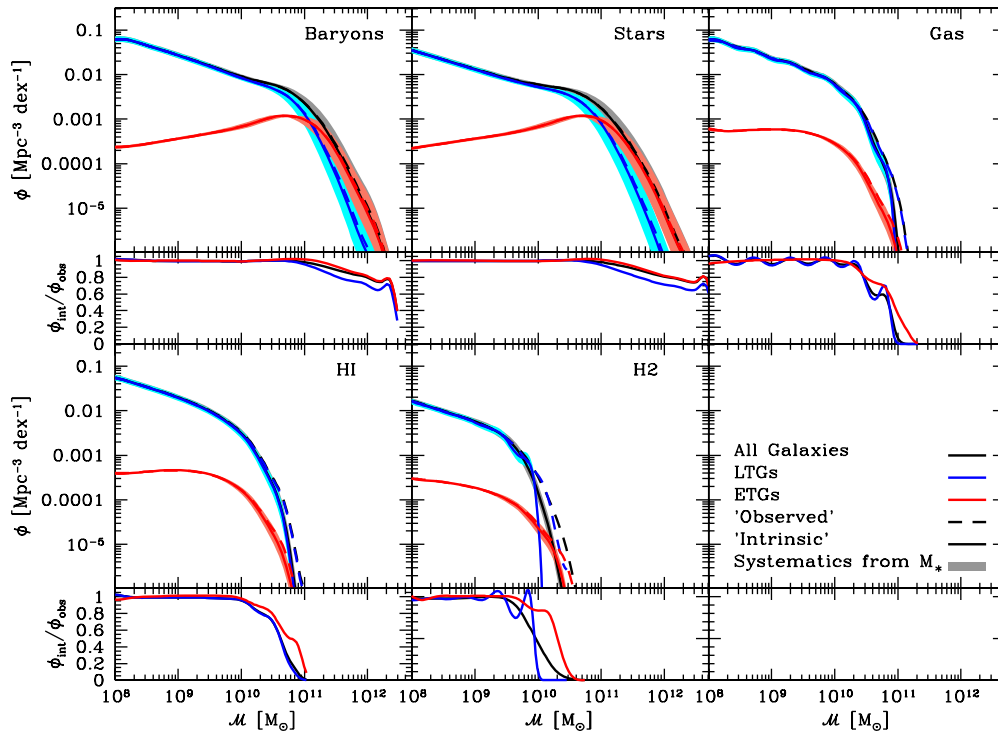### 5.1. The impact of random and systematics errors

When deriving stellar, HI and $H_2$ masses, there are two sources of errors that will inevitably propagate over the MFs: the random and systematic errors. In this Section, we discuss the impact of both sources of errors on our results.

#### 5.1.1. Random errors

The estimation of masses from both photometric and radio observations is subject to random errors. Here, we determine their impact on our resulting MFs. For simplicity, we assume that random errors follow lognormal distributions with a constant dispersion and independent of galaxy morphology. For the stellar masses, we assume a dispersion of $\sigma = 0.1$ dex following Behroozi, Conroy, & Wechsler (2010), Mendel et al. (2014), and Rodríguez-Puebla et al. (2017). For HI masses, $\sigma = 0.14$ dex, and for $H_2$ masses, $\sigma = 0.22$ dex, following Calette et al. (2018, and more references therein). As for the gas and baryonic masses, we assume errors of respectively $\sigma = 0.14$ dex and $\sigma = 0.1$ dex as they are dominated by HI and $M_*$ components, respectively, especially at high masses, where random errors have a larger impact. Thus, our 'observational' MFs[s] are the result of the convolution of the distribution of random errors and the respective intrinsic MFs. That is, our 'observational' MFs are given by $\phi_{obs} = \mathcal{G} * \phi_{int}$, where the symbol $*$ denotes the convolution operation, $\mathcal{G}$ is the distribution of random errors, and $\phi_{int}$ is the intrinsic MF. For more details, the reader is referred to Appendix D. There we describe our numerical algorithm for deconvolving the intrinsic MF, $\phi_{int}$.

In Figure 10, we reproduce with dashed lines the 'observational' MFs derived in Section 4.4. Their corresponding intrinsic MFs are shown with solid lines. In the same figure, we present the ratios $\phi_{obs}/\phi_{int}$ to show the effect of the deconvolution. The effect of deconvolving from random errors is small at low-intermediate masses but it increases at the massive end since the MFs are steeper (Eddington 1940). This is simply because the convolution depends on the logarithmic slope of the intrinsic MF (e.g., Cattaneo et al.

---

[s]In the preceding sections, we omit to use the term 'observational' MFs to avoid confusion about our methodology. Here, we use this term to refer that our determinations, similar to direct measurements of the MFs from galaxy surveys, suffer from random errors.

**Figure 10.** Impact of random and systematics errors in the baryonic, stellar, cold gas, atomic, and molecular gas MFs for all and separately for early- and late-type galaxies. The dashed lines show the 'observational' MF from Section 4.4, while the solid lines show the MF after deconvolving from random errors, i.e., the intrinsic MFs. Systematic errors are shown with the shaded areas. While the impact of random errors affects notably the total cold, atomic, and molecular gas MFs, the impact of systematic uncertainty on $M_*$ is apparently marginal on them. However, the systematic uncertainties on $M_*$ are noticeable in the stellar and baryonic MFs.

2008); the steeper the slope, the larger the effect on the MFs. This is also the reason why we observe a lower impact in the baryonic and stellar MFs compared to the HI, $H_2$, and cold gas MFs; the latter fall steeply at the high-mass end. For example, the intrinsic HI MF is a factor of $\sim 4$ lower than the 'observational' one at $M_{HI} \sim 6 \times 10^{10}$ $M_\odot$, while the intrinsic $H_2$ MF is an order of magnitude lower than the 'observational' MF at $M_{H2} \sim 2 \times 10^{10}$ $M_\odot$. The intrinsic gas MF is an order of magnitude lower than the 'observational' MF at $M_{gas} \sim 10^{11}$ $M_\odot$. Note that for the HI, $H_2$, and cold gas MFs the impact of random errors is more noticeable in late-type galaxies than in early-type ones.

### 5.1.2. Systematic errors

In addition to random errors, systematic errors have an impact when determining the MFs. The IMF is one of the most important sources of systematic errors for the GSMF. In this paper, we assumed an universal IMF given by the Chabrier (2003) function. While there is much debate on the IMF (see e.g., Bastian, Covey, & Meyer 2010; Conroy et al. 2013; Bernardi et al. 2018), exploring the different alternatives is beyond the scope of this paper.

The stellar masses are calculated typically using colour-dependent mass-to-light ratios based on results from stellar population synthesis (SPS) models (for a recent review, see Conroy 2013). Thus, the calculated stellar masses depend on the used SPS model. This introduces a systematic uncertainty in $M_*$. Indeed, systematics in $M_*$ from SPS can be as large as $\sim 0.25$ dex, see for example, Pérez-González et al. (2008); Muzzin et al. (2009); Moustakas et al. (2013); Rodríguez-Puebla et al. (2017) and references therein. Recently, Bernardi et al. (2017) showed that systematics from SPS introduces errors that are as large as $\sim 0.5$

dex in the normalisation of the GSMF at the high-mass end. In Appendix A, we found similar differences using various recipes of colour-dependent mass-to-light ratios. While in this paper, we calculate five different stellar masses for every galaxy and decided to use the geometric mean of the five as our fiducial definition of $M_*$, the above inevitable introduces the question of *which stellar mass definition shall we use when deriving our MFs*. Additionally, Bernardi et al. (2017) determined that systematics in photometry are of $\sim 0.1$ dex. In order to quantify the impact of stellar populations in our MFs, in Appendix A, we noted that a constant shift of $\sim \pm 0.15$ dex in the stellar mass axis reproduces systematic errors in the GSMF. In addition, Figure 1 shows that the same shift in the stellar mass axis could also explain differences from photometry. Thus, hereafter, we will use a shift of $\pm 0.15$ dex in the stellar mass axis as our fiducial model for systematic errors in the GSMF. Note that we are assuming that this shift will be independent of morphology and we are ignoring systematic errors in the atomic and molecular gas components.

Figure 10 shows the impact of systematic errors from SPS models and photometry as the shaded areas around their corresponding $\phi_{int}$ (solid lines). The effects of systematics is non-negligible at the massive end of the stellar and baryonic MFs; we observe differences up to $\sim 0.6$ dex in their normalisations. This is approximately the same both for early- and late-type galaxies. The impact of systematic errors in the gas, HI, and $H_2$ MFs is marginal; we notice a shift in their normalisations of $\sim 0.07$ at their low-mass ends but they increases respectively to $\sim 0.4$, $\sim 0.4$, and $\sim 0.3$ dex at their massive ends. The above is due to the steeper slopes observed at the high-mass end from these MFs. In conclusion, the impact of systematic uncertainties in $M_*$ is only marginal for the derived

HI, $H_2$, and cold gas MFs, making our results robust against this source of uncertainties.

As mentioned in the Introduction, the value of deriving robust MFs is that they can be used as key tools for constraining the processes that govern the evolution of the galaxies. However, using direct measurements from observations to constrain galaxy formation models is not trivial due to random and systematic errors, as discussed here. We end this subsection by emphasising the importance of deconvolving from random errors and understanding the impact of systematic errors when reporting results on galaxy demographics.

## 6. Summary and conclusions

We present a self-consistent empirical approach that unifies local galaxy gas-to-stellar mass correlations and the MFs of galaxies traced by their different baryonic components. We make available a PYTHON code that displays tables and figures with all the relevant statistical distributions and correlations discussed in this paper.[t] Next, we summarise our main results which can be used for comparing with theoretical predictions or as input for modelling galaxy mock catalogues:

- *Conditional probability distribution functions (CPDFs):* Section 2.2 presents the functional forms for the HI and $H_2$ mass conditional distributions given $M_*$ (the CPDFs), which are described by Equations (7)–(12). Our best-fit parameters to the empirical information presented in Paper I are listed in Table 3, while Figures 3 and 4 show the data with their corresponding best fits in various stellar mass bins. Theoretical predictions for the HI, $H_2$, and cold gas CPDFs can be confronted with our empirically constrained distributions, for all galaxies as well as for early and late types in case the morphological classifications are available. If these predictions are limited in stellar and/or gas masses, then our (analytical) HI-CPDFs and $H_2$-CPDFs and their moments can be easily calculated over the same mass ranges as the theoretical predictions for a comparison. The HI- and $H_2$-CPDFs combined with the GSMF allowed us to calculate the respective bivariate mass distributions for all the galaxy population as plotted in Figures 6 and 7.
- *Moments of the CPDFs:* The (analytical) HI- and $H_2$-CPDFs contain the information about any moment of the distributions. Figure 4 (see also Figures 6 and 7) shows the $\langle \log M_j \rangle + - \log M_*$ relationships, with $j =$ HI, $H_2$, for early- and late-type galaxies as well as for all galaxies. In addition, we present these relationships using the arithmetic mean, $\langle M_j \rangle$. As expected, these relationships lie above from those calculated with the logarithmic mean, $\langle \log M_j \rangle$. Moreover standard deviations can vary significantly if they are computed with respect to the arithmetic or logarithmic mean, which also depends on the shape of the distributions. Other statistical measures that can be used to characterise the population distributions are medians and percentiles, for example. As mentioned above, any statistical quantity can be computed with our CPDFs and confronted with both theoretical and/or observational results.
- *The Galaxy Stellar Mass Functions:* In Section 3.1, we determined the GSMF from the SDSS DR7 based on the photometric catalogue from Meert et al. (2015) and (2016) for masses above

$M_* = 10^9$ $M_\odot$. For masses down to $\sim 3 \times 10^7 M_\odot$, we used the low-z SDSS DR4 (Blanton et al. 2005a; b), and corrected it from surface brightness incompleteness and fluctuations due to large-scale structures. We determined also the fractions of early- and late-type galaxies using the SDSS DR7 morphological classification of Huertas-Company et al. (2011). Stellar masses were derived from five colour-dependent mass-to-light ratios. We used as our fiducial definition the geometric mean of these five stellar masses derived for each galaxy. We also determined the impact of systematic errors in $M_*$ due to mass-to-light ratio uncertainties in our MFs.
- *Calculated MFs:* Section 4.4 presents the results of calculating with our approach the MFs for atomic, molecular, cold gas, and baryons for early- and late-type galaxies, as well as for all galaxies. As discussed in Section 5.1, random errors in mass determinations artificially decrease the slope of the 'observational' MFs, an effect that affects especially the high-mass end, and that would lead to incorrect conclusions when comparing to theoretical predictions. Figure 10 presents our MFs deconvolved from random errors, that is, the *intrinsic* MFs, for different baryon matter components, and separately for early- and late-type galaxies. In the same section, we studied the effects on the MFs from systematic errors in $M_*$, also shown in Figure 10. In Appendix A and Figure 1, we showed explicitly that systematic errors in the GSMF due to mass-to-luminosity ratios and photometric uncertainties are well represented by a shift in the $M_*$-axis of $\pm 0.15$ dex. The effect of random errors in the baryonic MF is of the same order, while for the gas MFs the propagated systematic errors in $M_*$ have a negligible effect. Note that our MFs are complete only above a given mass limit, $\sim 3 \times 10^7$ $M_\odot$ for the GSMF, $\sim 10^7$ $M_\odot$ for the $H_2$ MF, and $\sim 10^8$ $M_\odot$ for the HI, cold gas, and baryonic MFs.

From the results summarised above, we highlight the following conclusions:

- The low-mass slope of our GSMF, corrected for surface brightness incompleteness, is $\alpha \approx -1.4$, consistent with recent determinations based on the deeper surveys such as GAMA (Wright et al. 2017), and estimations based on the search of low surface brightness galaxies from core-collapse supernovae (Sedgwick et al. 2019). The slope for the high-mass end is shallower than previous determination most likely as the result of the new photometric catalogue employed in this paper (Meert et al. 2015). Similar results have been reported in Bernardi et al. (2017).
- The total GSMF is well fitted by a function composed of a sub-exponential Schechter function and a double power law function. This fitting model has an error of less than $\sim 2\%$ in the mass range $2 \times 10^9 - 5 \times 10^{11}$ $M_\odot$. At the smallest and largest masses, the deviations increase to values above $\sim 20\%$. In contrast, the commonly employed double Schechter function model performs considerably worse.
- Systematic errors due to SPS models, that affect results on mass-to-light ratios, introduce a systematic effect on the normalisation of the GSMF, especially at the massive end. We find differences between $\sim 0.5 - 1$ dex, consistent with the result discussed in Bernardi et al. (2017).
- The HI, $H_2$, and cold gas MFs are mostly dominated by late-type galaxies. In general, we notice that our HI MF is in good agreement with previous determinations from blind surveys. Similarly, the $H_2$ MF is consistent with previous determinations

based on CO follow-up optically selected samples. When we compare to the HIPASS and ATLAS 3D surveys for early-type galaxies, our HI MF is consistent with those observations. However, our $H_2$ MF for early-type galaxies is in tension at the low-mass side with the MF derived from the ATLAS 3D survey.

- Our 'observational' MFs were deconvolved from random errors to obtain the intrinsic MFs. The effect of random errors is small at the low-mass end but larger at the high-mass end of our MFs. This is because the convolution depends on the logarithmic slope of the intrinsic MFs. Because the baryonic and stellar MFs are shallower at the massive end, the effects are relatively small, but the atomic, molecular, and cold gas MFs have steeper slopes resulting in a larger effect.

- While for the stellar (and hence baryonic) MF systematic errors due to mass-to-light ratio uncertainties introduce a non-negligible effect, especially at the high-mass end, for the atomic, molecular, and gas MFs, the effects of systematics are small. We thus conclude that our determinations for the gas MFs are robust against systematic errors in the the $M_*$ determination.

- We determined the $z \sim 0$ cosmic densities of HI, $H_2$, cold gas, stars, and baryons locked in galaxies calculated from the respective MFs. Our results are in good agreement with previous determinations from different local censuses. Most of the atomic and molecular H gas is in late-type galaxies, $\sim$96% of the mass density, while this fraction decreases to $\sim$70% and $\sim$65% for baryons and stars. We find that the fraction of HI and $H_2$ in galaxies with respect to the universal baryon fraction is respectively $\sim$1% and $\sim$0.2%, while the respective fractions for mass in stars is $\sim$4%. Baryons in galaxies (the ionised and hot gas were not included) are $\sim$5.4% of the universal baryon fraction.

- Based on the values reported in the literature for the local CSFR of star-forming (late-type) galaxies, we estimated the cosmic $H_2$ and total gas depletion times of late-type galaxies. These timescales, $t_{dep}(H_2) \approx 1.3$ Gyr and $\bar{t}_{dep,L}(gas) \approx 10.14$ Gyr, respectively, imply that galaxies, on average, are inefficient to convert their molecular gas into stars, and are inefficient to transform their atomic gas into molecular gas. The depletion time for the total neutral hydrogen is $\bar{t}_{dep,L}(H) = \bar{t}_{dep,L}(gas)/1.4 \sim 7.25$ Gyrs. On the other hand, the average cosmic SF timescale (the inverse of the cosmic sSFR) is $\bar{t}_{SF,L} \approx 20.3$ Gyrs, which implies that the ratio $\bar{t}_{dep,L}(H)/\bar{t}_{SF,L} = 0.38$. This shows that the gas reservoir of late-type galaxies has not yet been dramatically consumed by star formation.

Here, we provided a statistical description for calculating any moment to characterise the gas-to-stellar mass correlations, the HI- and $H_2$-stellar mass bivariate distributions as well as all the respective MFs. One of our motivations for this paper is to provide the community with a full self-consistent phenomenological description of the local galaxy population for various properties and divided into the two main morphological types in order to be confronted with theoretical results, both from semi-analytical models and cosmological hydrodynamical simulations. The next generation of sensitive radio telescopes will be able to survey large samples of extragalactic sources in HI and $H_2$ gas, something that is a common practice with current optical surveys. Thus, robust and unbiased bivariate distributions and MFs of HI and $H_2$ gas over large mass ranges will be routinely derived in the future along with the relationships of the gas contents with their optical/IR properties. Preparatory to that, and to pave the road to these surveys, studies based on radio follow-up observations of

(relative small) optically selected galaxy samples provide valuable information that can be used for the gas demographics of galaxies. In this work, we have exploited the results from many of these studies, and by means of the conditional (or bivariate) approach we were able to derive the abundances of local galaxies as traced by different baryonic components and separated into the two main groups of galaxies, early, and late types.

The present work is the second paper of a series. In Paper I, we derived the CPDFs of HI and $H_2$ as a function of $M_*$, separately for early- and late-type galaxies, for an extensive compilation and homogenisation of radio data from the literature. In the present work, we made extensive use of these data. In the future, we will use the MFs derived here to extend the galaxy–halo connection for different baryonic components, and we will show that not only the HI and $H_2$ MFs derived here are in good agreement with radio blind or optically selected surveys but also with the observed galaxy spatial clustering as a function of HI gas mass.

## References

Andreani, P., Boselli, A., Ciesla, L., Vio, R., Cortese, L., Buat, V., & Miyamoto, Y. 2018, A&A, 617, A33

Aver, E., Olive, K. A., & Skillman, E. D. 2015, JCAP, 2015, 011

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681

Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., & Budavari, T. 2006, MNRAS, 373, 469

Baldry, I. K., Glazebrook, K., & Driver, S. P. 2008, MNRAS, 388, 945

Baldry, I. K., et al. 2012, MNRAS, 421, 621

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339

Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Bernardi, M., Shankar, F., Hyde, J. B., Mei, S., Marulli, F., & Sheth, R. K. 2010, MNRAS, 404, 2087

Bernardi, M., Meert, A., Sheth, R. K., Vikram, V., Huertas-Company, M., Mei, S., & Shankar, F. 2013, MNRAS, 436, 697

Bernardi, M., Meert, A., Sheth, R. K., Huertas-Company, M., Maraston, C., Shankar, F., & Vikram, V. 2016, MNRAS, 455, 4122

Bernardi, M., Meert, A., Sheth, R. K., Fischer, J.-L., Huertas-Company, M., Maraston, C., Shankar, F., & Vikram, V. 2017, MNRAS, 467, 2217

Bernardi, M., Sheth, R. K., Dominguez-Sanchez, H., Fischer, J.-L., Chae, K.-H., Huertas-Company, M., & Shankar, F. 2018, MNRAS, 477, 2560

Bertin E. & Arnouts S. 1996, A&AS, 177, 393

Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Blanton, M. R., et al. 2005a, AJ, 129, 2562

Blanton, M. R., Lupton, R. H., Schlegel, D. J., Strauss, M. A., Brinkmann, J., Fukugita, M., & Loveday, J. 2005b, ApJ, 631, 208

Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, AJ, 142, 31

Braun R. 2012, ApJ, 749, 87

Buckley, J., & James, I. 1979, Biometrika, 66, 429

Butcher, Z., Schneider, S., van Driel, W., & Lehnert, M. D. 2018, A&A, 619, A89

Calette, A. R., Avila-Reese, V., Rodrguez-Puebla, A., Hernández-Toledo, H., & Papastergis, E. 2018, RMxAA, 54, 443

Cannon, J. M., et al. 2015, AJ, 149, 72

Catinella, B., et al. 2013, MNRAS, 436, 34

Cattaneo, A., Dekel, A., Faber, S. M., & Guiderdoni, B. 2008, MNRAS, 389, 567

Chabrier, G. 2003, PASP, 115, 763

Chilingarian, I. V., Melchior, A.-L., & Zolotukhin, I. Y. 2010, MNRAS, 405, 1409

Conroy, C. 2013, ARA&A, 51, 393

Conroy, C., Dutton, A. A., Graves, G. J., Mendel, J. T., & van Dokkum, P. G. 2013, ApJ, 776, L26

Conselice, C. J., Wilkinson, A., Duncan, K., & Mortlock, A. 2016, ApJ, 830, 83

Cooke, R. J., Pettini, M., & Steidel, C. C. 2018, ApJ, 855, 102

Croton, D. J., et al. 2016, ApJS, 222, 22

D'Souza, R., Vegetti, S., & Kauffmann, G. 2015, MNRAS, 454, 4027

Davé, R., Anglés-Alcázar, D., Narayanan, D., Li, Q., Rafieferantsoa, M. H., & Appleby, S. 2019, MNRAS, 486, 2827

Delhaize, J., Meyer, M. J., Staveley-Smith, L., & Boyle, B. J. 2013, MNRAS, 433, 1398

Diemer, B., et al. 2018, ApJS, 238, 33

Diemer, B., et al. 2019, arXiv e-prints,

Dragomir, R., Rodrguez-Puebla, A., Primack, J. R., & Lee, C. T., 2018, MNRAS, 476, 741

Eddington Sir, A. S. 1940, MNRAS, 100, 354

Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 232, 431

Feigelson, E. D., & Nelson, P. I. 1985, ApJ, 293, 192

Fliri, J., & Trujillo, I. 2016, MNRAS, 456, 1359

Frenk, C. S., & White, S. D. M. 2012, Annalen der Physik, 524, 507

Frieman, J. A., et al. 2008, AJ, 135, 338

Fukugita, M., & Peebles, P. J. E. 2004, ApJ, 616, 643

Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518

Fukugita, M., et al. 2007, AJ, 134, 579

Haynes, M. P., et al. 2011, AJ, 142, 170

Haynes, M. P., et al. 2018, ApJ, 861, 49

He, Y. Q., Xia, X. Y., Hao, C. N., Jing, Y. P., Mao, S., & Li, C., 2013, ApJ, 773, 37

Henriques, B. M. B., White, S. D. M., Lilly, S. J., Bell, E. F., Bluck, A. F. L., & Terrazas, B. A. 2019, MNRAS, 485, 3446

Hernández-Toledo, H. M., Vázquez-Mata, J. A., Martínez-Vázquez, L. A., Choi, Y.-Y., & Park, C. 2010, AJ, 139, 2525

Hirschmann, M., Dolag, K., Saro, A., Bachmann, L., Borgani, S., & Burkert, A. 2014, MNRAS, 442, 2304

Hoppmann, L., Staveley-Smith, L., Freudling, W., Zwaan, M. A., Minchin, R. F., & Calabretta, M. R. 2015, MNRAS, 452, 3726

Hu, W., et al. 2019, MNRAS, 489, 1619

Huertas-Company, M., Aguerri, J. A. L., Bernardi, M., Mei, S., & Sánchez Almeida, J. 2011, A&A, 525, A157

Hyde, J. B., & Bernardi, M. 2009, MNRAS, 393, 1171

Jones, M. G., Haynes, M. P., Giovanelli, R., & Moorman, C. 2018, MNRAS, 477, 2

Kaplan, E. L., & Meier, P. 1958, Journal of the American Statistical Association, 53, pp. 457

Kelvin, L. S., et al. 2014, MNRAS, 444, 1647

Kennicutt Jr., R. C. 1998, ApJ, 498, 541

Keres, D., Yun, M. S., & Young, J. S. 2003, ApJ, 582, 659

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S., 2016, MNRAS, 457, 4340

Koribalski, B. S., et al. 2004, AJ, 128, 16

Kovac, K., Oosterloo, T. A., & van der Hulst, J. M. 2005, in IAU Colloq. 198: Near-Fields Cosmology with Dwarf Elliptical Galaxies, ed. H. Jerjen & B. Binggeli, 351–354 (arXiv:astro-ph/0508072), doi:10.1017/S1743921305004047

Kravtsov, A., Vikhlinin, A., & Meshscheryakov, A. 2014, preprint, arXiv:1401.7329

Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687

Lada, C. J., Forbrich, J., Lombardi, M., & Alves, J. F. 2012, ApJ, 745, 190

Lagos C. d. P., Davis, T. A., Lacey, C. G., Zwaan, M. A., Baugh, C. M., Gonzalez-Perez, V., & Padilla, N. D. 2014, MNRAS, 443, 1002

Lagos C. d. P., et al. 2015, MNRAS, 452, 3815

Lagos C. d. P., Tobar, R. J., Robotham, A. S. G., Obreschkow, D., Mitchell, P. D., Power, C., & Elahi, P. J. 2018, MNRAS, 481, 3573

Lemonias, J. J., Schiminovich, D., Catinella, B., Heckman, T. M., & Moran, S. M. 2013, ApJ, 776, 74

Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2782

Leroy, A. K., et al. 2013, AJ, 146, 19

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Maddox, N., Hess, K. M., Obreschkow, D., Jarvis, M. J., & Blyth, S. L. 2015, MNRAS, 447, 1610

Maeda, F., Ohta, K., & Seko, A. 2017, ApJ, 835, 120

Martin, A. M., Papastergis, E., Giovanelli, R., Haynes, M. P., Springob, C. M., & Stierwalt, S. 2010, ApJ, 723, 1359

Masters, K. L., et al. 2010a, MNRAS, 404, 792

Masters, K. L., et al. 2010b, MNRAS, 405, 783

Meert, A., Vikram, V., & Bernardi M. 2013, MNRAS, 433, 1344

Meert, A., Vikram, V., & Bernardi, M. 2015, MNRAS, 446, 3943

Meert, A., Vikram, V., & Bernardi, M. 2016, MNRAS, 455, 2440

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3

Meyer, M. J., et al. 2004, MNRAS, 350, 1195

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Moffett, A. J., et al. 2016a, MNRAS, 457, 1308

Moffett, A. J., et al. 2016b, MNRAS, 462, 4336

Moustakas, J., et al. 2013, ApJ, 767, 50

Muzzin, A., Marchesini, D., van Dokkum, P. G., Labbé, I., Kriek, M., & Franx, M. 2009, ApJ, 701, 1839

Nair, P. B., & Abraham, R. G. 2010, ApJS, 186, 427

Obreschkow, D., & Rawlings, S. 2009, MNRAS, 394, 1857

Obreschkow, D., Murray, S. G., Robotham, A. S. G., & Westmeier, T., 2018, MNRAS, 474, 5500

Papastergis, E., Cattaneo, A., Huang, S., Giovanelli, R., & Haynes, M. P. 2012, ApJ, 759, 138

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Peng, Y.-j., et al. 2010, ApJ, 721, 193

Peng, Y.-j., Lilly, S. J., Renzini, A., & Carollo, M. 2012, ApJ, 757, 4

Pérez-González, P. G., et al. 2008, ApJ, 675, 234

Pillepich, A., et al. 2018, MNRAS, 473, 4077

Planck Collaboration et al. 2016, A&A, 594, A13

Planck Collaboration et al. 2018, arXiv e-prints, p.arXiv:1807.06209

Popping, G., et al. 2019, arXiv e-prints,

Read, J. I., & Trentham, N. 2005, Philosophical Transactions of the Royal Society of London Series A, 363

Rodríguez-Puebla, A., Avila-Reese, V., & Drory, N. 2013, ApJ, 767, 92

Rodríguez-Puebla, A., Behroozi, P., Primack, J., Klypin, A., Lee, C., & Hellinger, D. 2016, MNRAS, 462, 893

Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, MNRAS, 470, 651

Romeo, A. B. 2020, MNRAS, 491, 4843

Saintonge, A., et al. 2011, MNRAS, 415, 32

Saintonge, A., et al. 2017, ApJS, 233, 22

Sánchez, S. F., et al. 2019, MNRAS, 482, 1557

Schaye, J., et al. 2015, MNRAS, 446, 521

Schmidt, M. 1968, ApJ, 151, 393

Sedgwick, T. M., Baldry, I. K., James, P. A., & Kelvin, L. S. 2019, MNRAS, 484, 5278

Serra, P., et al. 2012, MNRAS, 422, 1835

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, ApJS, 160, 149

Takeuchi, T. T. 2010, MNRAS, 406, 1830

Takeuchi, T. T., Sakurai, A., Yuan, F.-T., Buat, V., & Burgarella, D. 2013, Earth, Planets, and Space, 65, 281

Taylor, E. N., et al. 2011, MNRAS, 418, 1587

Tempel, E., et al. 2014, A&A, 566, A1

Thanjavur, K., Simard, L., Bluck, A. F. L., & Mendel, T. 2016, MNRAS, 459, 44

Tonry, J. L., Blakeslee, J. P., Ajhar, E. A., & Dressler, A. 2000, ApJ, 530, 625

Vikram, V., Wadadekar, Y., Kembhavi, A. K., & Vijayagovindan, G. V. 2010, MNRAS, 409, 1379

Vogelsberger, M., et al. 2014, MNRAS, 444, 1518

Willick, J. A., Courteau, S., Faber, S. M., Burstein, D., Dekel, A., & Strauss, M. A. 1997, ApJS, 109, 333

Wright, A. H., et al. 2017, MNRAS, 470, 283

Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., & Han, J. 2012, ApJ, 752, 41

Young, J. S., et al. 1995, ApJS, 98, 219

Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Popping, G., & Davé, R. 2019, MNRAS, 483, 2983

Zibetti, S., Charlot, S., & Rix, H.-W. 2009, MNRAS, 400, 1181

Zwaan, M. A., Briggs, F. H., Sprayberry, D., & Sorar, E. 1997, ApJ, 490, 173

Zwaan, M. A., et al. 2003, AJ, 125, 2842

Zwaan, M. A., Meyer, M. J., Staveley-Smith, L., & Webster, R. L. 2005, MNRAS, 359, L30

van Driel, W., et al. 2016, A&A, 595, A118

van den Bosch, F. C., Aquino, D., Yang, X., Mo, H. J., Pasquali, A., McIntosh, D. H., Weinmann, S. M., & Kang, X. 2008, MNRAS, 387, 79

## A. Derivation of the SDSS DR7 GSMF

In this paper, we derive the GSMF from a spectroscopic sample of 670,722 galaxies from the SDSS DR7 based on the photometric estimates of the apparent brightnesses in the $g$, $r$, and $i$ band from Meert et al. (2015) and (2016). In those papers, the authors selected galaxies with extinction-corrected $r-$band Petrosian magnitude between magnitude 14 and 17.77 to derive de Vacouleurs, Sérsic, de Vacoulers+Exponential, and Sérsic+Exponential fits to the observed surface brightness profiles of each galaxies in their SDSS DR7 catalogue. Surface brightness profiles were obtained via the PyMorph pipeline (Vikram et al. 2010; Meert et al. 2013). PyMorph is a python software that uses Sextractor (Bertin & Arnouts 1996) and Galfit (Peng et al. 2002) to fit both one- and two-component to the seeing convolved surface brightness profiles from the spectroscopic sample of SDSS DR7 galaxies. PyMorph has been extensively tested in (Meert et al. 2013, see also, Meert et al. 2015) showning that the algorithm does not suffer from the sky subtraction problems that has been detected in previous studies based on the SDSS, in particular in crowded fields[u].

We estimate the GSMF at the redshift interval between $z = 0.005$ and $z = 0.2$ using the standard $1/V_{max}$ method (Schmidt 1968):

$$\phi_*(M_*) = \frac{1}{\Delta \log M_*} \sum_{i=1}^{N} \frac{\omega_i(\log M_* \pm \Delta \log M_*)}{V_{max,i}}, \qquad (A1)$$

where $\omega_i$ is a weight factor correction that depends on the position in the sky for galaxies within the interval $\log M_* \pm \Delta \log M_*/2$, following Bernardi et al. (see also 2010) we assume that $\omega_i = 1.1$; and

$$V_{max,i} = \int_{\Omega} \int_{z_{l,i}}^{z_{u,i}} \frac{d^2 V_c}{dz d\Omega} dz d\Omega. \qquad (A2)$$

---

[u]Recently, various others groups have also improved the determinations of galaxies' surface brightness profiles based on the SDSS by the improving the survey photometry, especially due to sky subtraction problems in crowded fields, (see e.g., Simard et al. 2011; D'Souza et al. 2015, and more reference therein). While in this paper we opt to use the photometric catalogue from Meert et al. (2015) and (2016), Bernardi et al. (2017) showed that, after a careful comparison, most of those studies agree up to 0.1 dex. Thus, using the photometry derived by other groups will not change significantly our results.

We denote the solid angle of the SDSS DR7 with $\Omega$, while $V_c$ refers to the comoving volume enclosed within the redshift interval $[z_{l,i}, z_{u,i}]$. The redshift limits are defined as $z_{l,i} = \max(0.005, z_{min,i})$ and $z_{u,i} = \min(z_{max,i}, 0.2)$; where $z_{min,i}$ and $z_{max,i}$ are, respectively, the minimum and maximum at which each galaxy can be observed in the SDSS DR7 sample. We estimate $z_{max,i}$ for every galaxy in the sample by solving iteratively the distance modulus equation:

$$m_{lim,r} - M_{r,i}^{0.0} = 5 \log D_{L,i}(z_{max,i}) + 25 \qquad (A3)$$

$$+ K_{gr,i}(z_{max,i}) - E_{r,i}(z_{max,i}), \qquad (A4)$$

where $M_{r,i}^{0.0}$ is the Petrosian magnitude K+E-corrected at a rest frame $z = 0$, $K_{gr,i}(z)$ is the $K$-correction (see Appendix B) and $E_{r,i} = 1.1z$ (following Dragomir et al. 2018) for the $i$th galaxy in the sample. For the completeness limits, we use the limiting apparent magnitude in the $r$-band of $m_{lim,r} = 17.77$. Similarly, we estimate $z_{min,i}$ by solving iteratively the distance modulus equation but this time using the limiting apparent magnitude $m_{lim,r} = 14$.

Errors are estimated using the jackknife technique by diving the galaxy sample into $n = 300$ subsamples of approximately equal size and estimating a $\phi_{*,i}(M_*)$ each time. Thus, errors are then given by:

$$\sigma^2 = \frac{n-1}{n} \sum_{i=1}^{n} \left( \phi_{*,i} - \langle \phi_* \rangle \right)^2, \qquad (A5)$$

with $\langle \phi_* \rangle$ as the average of the ensemble.

Stellar masses were derived from several colour-dependent mass-to-light ratios as listed below:

$$M_* = \begin{cases} \Upsilon_r^{B03}(g-r) \cdot L_r \text{ Bell et al. (2003)} \\ \Upsilon_i^{B03}(g-i) \cdot L_i \text{ Bell et al. (2003)} \\ \Upsilon_r^{Z09}(g-r) \cdot L_r \text{ Zibetti, Charlot, \& Rix et al. (2009)}, \\ \Upsilon_i^{Z09}(g-i) \cdot L_i \text{ Zibetti et al. (2009)} \\ \Upsilon_i^{T11}(g-i) \cdot L_i \text{ Taylor et al. (2011)} \end{cases}$$

$$(A6)$$

and we define our fiducial $M_*$ as the geometrical mean of all the determinations in 34:

$$M_* = [M_*(\Upsilon_r^{B03}) \times M_*(\Upsilon_i^{B03}) \times M_*(\Upsilon_r^{Z09}) \times \qquad (A7)$$

$$M_*(\Upsilon_i^{Z09}) \times M_*(\Upsilon_i^{T11})]^{1/5}$$

We use Sérsic apparent magnitudes to derive colours and magnitudes. We apply K+E-corrections at a rest frame $z = 0$. We use the values reported in Dragomir et al. (2018) for $g$ and $i$ bands given, respectively, by $E_g = 1.3 \times z$ and $E_i = 1.09 \times z$. K-corrections are discuss in Appendix B. We applied a shift of $-0.1$ dex to the resulting masses from the colour-dependent mass-to-light ratios of Bell et al. (2003) to be consistent with the Chabrier (2003) IMF adopted in this paper.

Figure 11 shows the resulting GSMFs as described above. The figure shows that using different recipes of deriving stellar masses yield to differences of $\sim 0.1$ dex at low masses and between $\sim 0.5 - 1$ dex at the high-mass end. This is consistent with the recent study by (Bernardi et al. 2017) which showed that differences in mass-to-light ratios introduce discrepancies in the GSMF around $\sim 0.5$ dex. As a fiducial estimation of the GSMF in this paper, we opt to utilise the geometric mean of all the masses derived based on the colour-dependent mass-to-light ratios listed
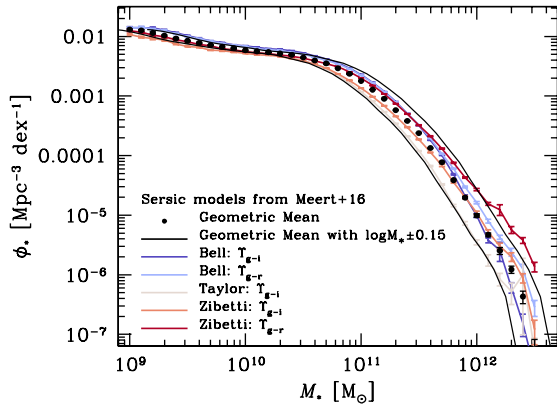
**Figure A1.** The GSMF from our six stellar mass definitions, Equation A6. Using different stellar masses yield to differences of ∼0.1 dex at low masses and as high as ∼1 dex at the high-mass end. In this paper, we opted to use as our fiducial GSMF as the one derived from the geometric mean of five different stellar masses. The solid lines show a shift of ±0.15 dex in the stellar mass axis of the our fiducial GSMF, note that it recovers systematics from mass-to-light ratios.

in Equation (34), filled circles with error bars. The black solid line shows a shift of ±0.15 dex in the stellar mass axis of our fiducial GSMF. Note that these shifts recover most of the differences observed due to systematics in mass-to-light ratios.

## B. K-corrections

Figure C1 shows the colour and redshift dependence of the K-corrections at a rest frame $z = 0$ for the $r$ (upper left), $g$ (upper right), and $i$ (bottom left) bands, shown as the dashed lines, from the NYU-VAGC SDSS DR7 and calculated from the K-CORRECT algorithm (v4_1_4 Blanton & Roweis 2007). In the same figure, the solid lines show the best fit according to the following relations:

$$K_j(z|\mathcal{C}) = \mathbf{Z}\mathbf{K}_j^{\mathcal{C}}\mathbf{C}, \tag{B1}$$

where $j$ denotes the $r$, $g$, and $i$ bands, while $\mathcal{C}$ denotes the uncorrected $g - r$ and $g - i$ galaxy colours. A similar approach has been done in Chilingarian, Melchior, & Zolotukhin (2010). The **C** and **Z** matrices are respectively given by

$$\mathbf{C} = \begin{pmatrix} 1 \\ C \\ C^2 \\ C^3 \\ C^4 \\ C^5 \end{pmatrix}, \tag{B2}$$

$$\mathbf{Z} = \begin{pmatrix} z & z^2 & z^3 & z^4 & z^5 \end{pmatrix}, \tag{B3}$$

while the $\mathbf{K}_j^{\mathcal{C}}$ matrices for the $r$, $g$, and $i$ bands are respectively given by Equations (B4), (B5), and (B6). Note that our K-corrections are polynomials of degree 5 in both colour and redshift and that in the above set of Equation $K_j(z=0|\mathcal{C})=0$.

$$K_r^{g-r} = \begin{pmatrix} 0.894302 & 2.32866 & -0.787673 & 0.324352 & -0.239774 & 0.0444971 \\ -15.5648 & 1.544 & -2.70992 & 3.42484 & -0.280197 & -0.0221534 \\ 49.7443 & -4.64543 & -8.72852 & 1.14138 & -1.76882 & 0.0702624 \\ -48.9173 & -4.95549 & 2.06966 & 14.5241 & -2.48092 & -0.322153 \\ 3.65716 & 21.3194 & -0.593275 & -6.04982 & -0.157727 & 0.731093 \end{pmatrix} \tag{B4}$$

$$K_g^{g-r} = \begin{pmatrix} 0.0786144 & 4.01535 & -0.883155 & 0.707471 & -2.05303 & 0.793141 \\ -6.81272 & 12.0599 & -10.7157 & 22.086 & -5.46384 & -1.34602 \\ -7.17353 & -52.5682 & -13.5845 & 11.2634 & -6.25812 & 2.61254 \\ 86.1835 & 96.7938 & -72.2792 & -1.44621 & -5.6531 & 9.09575 \\ -106.868 & -23.5461 & 101.815 & -43.5146 & 40.8195 & -21.677 \end{pmatrix} \tag{B5}$$

$$K_i^{g-i} = \begin{pmatrix} -3.01597 & 3.287 & -0.455067 & 0.426496 & -0.242669 & 0.0283777 \\ -1.11123 & -3.04641 & -5.2804 & 2.60911 & 0.134077 & -0.0813698 \\ 68.4078 & -14.6203 & -5.06879 & 0.904234 & -1.82776 & 0.47701 \\ -145.044 & 45.4714 & 8.75605 & 5.9425 & -1.32215 & -0.211679 \\ 59.2903 & -12.387 & -10.8653 & -1.84054 & 0.843326 & -0.0248045 \end{pmatrix} \tag{B6}$$

## C. GSMF for low-mass galaxies

### C.1. Surface brightness correction completeness

In this paper, we are interested in deriving the GSMF over a wide dynamical mass range, that is, from dwarf galaxies to massive elliptical galaxies. In Appendix A, we describe that based on the SDSS DR7 galaxy sample, we determined the GSMF for galaxies above $M_* \sim 10^9 M_\odot$. In this section, we determine the GSMF for galaxies above $M_* \sim 10^7 M_\odot$. Deriving the GSMF could be very challenging since the fraction of galaxies of missing galaxies due to surface brightness limits becomes very relevant at the faint end of the GSMF. Here, we follow a very simple statistical approach in order to quantify the number of galaxies missed due to surface brightness incompleteness limits as described in Blanton et al. (2005b). Our galaxy sample consist of a small volume ($0.0033 < z < 0.05$) carefully constructed to study very low-mass/luminosity galaxies from the SDSS NYU-VAGC with a total of 49968 galaxies (Blanton et al. 2005a; b)[v]. Hereafter, we will refer to this galaxy sample as the low-z SDSS.

Blanton et al. (2005b) estimated that the low-z SDSS galaxy sample has a completeness >70% for galaxies in the effective surface brightness range of $18 < \mu_{50,r} < 24$ mag arcsec$^{-2}$ and we consider galaxies only within this range. We assign to each galaxy a weight, $w_{\mu,j}$, which is a function of their central surface brightness and it takes into account the spectroscopic incompleteness ($1/w_{s,j}$), photometric incompleteness ($1/w_{p,j}$), and tiling catalogue incompleteness ($1/w_{t,j}$) in the sample. Thus, $w_{\mu,j} = w_{s,j} \times w_{p,j} \times w_{t,j}$. These weights were studied in detail in Blanton et al. (2005b) and provide the correlation between $w_{\mu,j}$ and effective surface brightness, $\mu_{50,r}$, in a tabulated form, see their Table 1. We use

[v] http://sdss.physics.nyu.edu/vagc/lowz.html
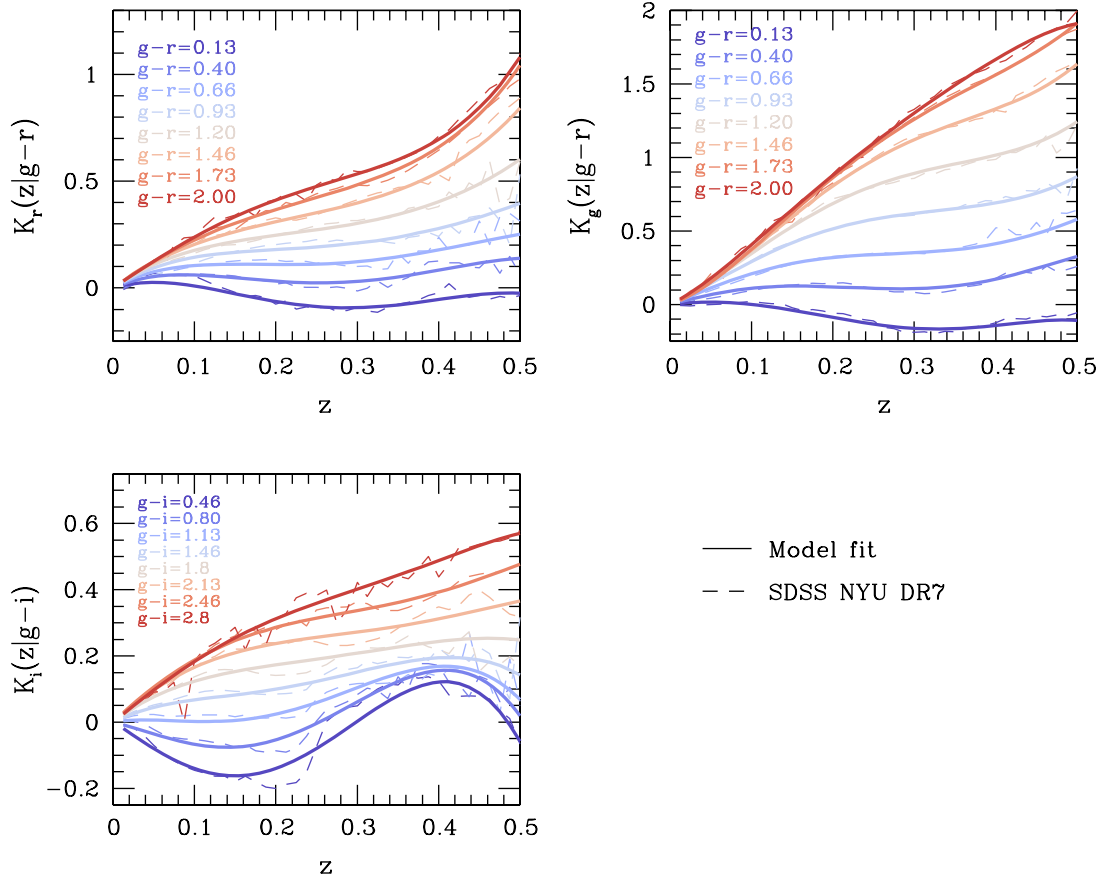
**Figure B1.** Colour and redshift dependence of the K-corrections at a rest frame $z = 0$ for the $r$ (upper left), $g$ (upper right), and $i$ (bottom left) bands from the K-CORRECT algorithm (Blanton & Roweis 2007), dashed lines. Solid lines show our best-fit models as described in the text.

cubic spline interpolations of this table in order to assign weight $w_{\mu,j}$ to each galaxy in the sample.

The next step in our programme is to estimate the number of missed galaxies brighter than $\mu_{50,r} = 24$ mag arcsec-2. To that end, we introduce a model for the distribution of $\mu_{50,r}$ as a function of $M_*$. We define the fraction of missing galaxies brighter than $\mu_{50,r} = 24$ mag arcsec-2 as a function of stellar mass as:

$$f_{\text{loss}}(M_*) = \frac{\sum_j N_{\text{obs}}(\mu_{50,j}|M_*)}{\sum_j N_{\text{real}}(\mu_{50,j}|M_*)}, \quad \text{(C1)}$$

where $N_{\text{real}}(\mu_{50,j}|M_*)$ and $N_{\text{obs}}(\mu_{50,j}|M_*)$ are the real and observed number galaxies with surface brightness (SB) between $\mu_{50,j} \pm d\mu_{50,j}$ and stellar masses between $\log M_* \pm d \log M_*/2$, respectively. Thus, our problem reduces to estimate $N_{\text{real}}$. Let us now define $P(\mu_{50,r}|M_*)$ as the conditional probability distribution of galaxies with SB $\mu_{50,r} \pm d\mu_{50,r}/2$ at a stellar mass bin $\log M_* \pm d \log M_*/2$. We calculate $P(\mu_{50,r}|M_*)$ directly from our galaxy sample by dividing it into stellar mass bins of 0.5 dex. This is done only for galaxies with Sérsic index $n_s \leq 2$ (galaxies with $n_s > 2$ are mostly of high SB, $\mu_{50,r} < 24$ mag arcsec$^{-2}$, so that it is not necessary to correct them for missing galaxies). For each stellar mass bin, we perform an extra binning of 0.05 dex in SB. We carry out the mentioned binning in $M_*$ and $\mu_{50,r}$ for each of the six different stellar mass estimators described above. As an example, the upper panel in Figure C2 show the distributions $P(\mu_{50,r}|M_*)$ for

one of our stellar mass estimators (empty circles). For each $M_*$ bin, we fit $P(\mu_{50,r}|M_*)$ assuming that it is described by a lognormal distribution,

$$P(\mu_{50,r}|M_*) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left[-\frac{\left(\mu_{50,r} - \langle\mu_{50,r}(M_*)\rangle\right)^2}{2\sigma_\mu^2}\right], \quad \text{(C2)}$$

where $\langle\mu_{50,r}(M_*)\rangle$ and $\sigma_\mu$ are the mean SB at a given stellar mass and the dispersion around it. We fit these two free parameters for each stellar mass bin. The best fits are plotted with solid lines. This operation is carried out for each of the stellar mass estimators used here. In the bottom panels of Figure C2, we show an example of the resulting best fits to observations as circles with error bars in the bivariate $(\mu_{50,r}, M_*)$ distribution plane (grey iso-contours) for three of our stellar mass estimators. The dashed lines show our SB magnitudes limits. Similarly to Baldry et al. (2008), we find that the relation between $\langle\mu_{50,r}\rangle$ and $\log M_*$ is linear for galaxies above $M_* \sim 10^9 M_\odot$[w] (filled circles) in the right panels of Figure C2. Departures from this linearity for galaxies below $M_* \sim 10^9 M_\odot$ (filled circles) is an indication that the relation between $\langle\mu_{50,r}\rangle$ and $\log M_*$ is affected by SB incompleteness. We fit the relationship between $\langle\mu_{50,r}\rangle$ and $\log M_*$ for galaxies above

---

[w]In fact, Baldry et al. (2008) found that the linearity holds above masses $M_* \sim 10^{8.5} M_\odot$. Here we apply the conservative value of $M_* \sim 10^9 M_\odot$. Nevertheless, we have found that using either Baldry et al. (2008) or our limit, the correction for SB is practically the same.
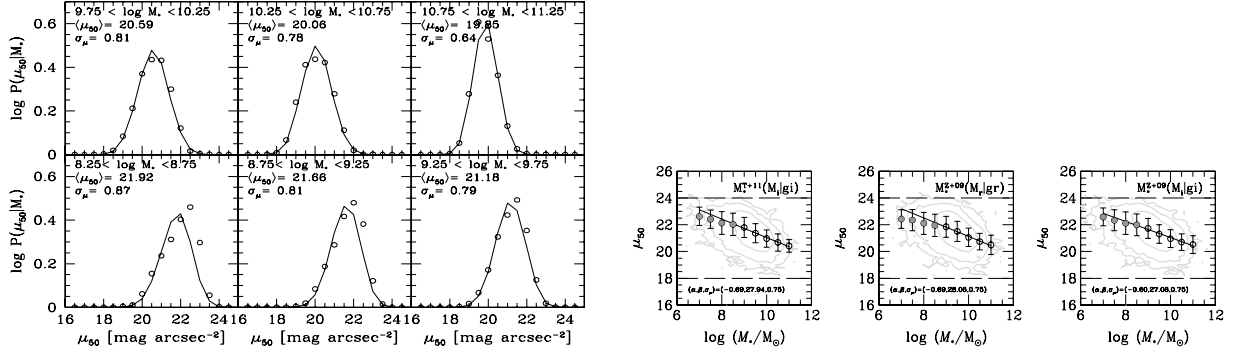
**Figure C1.** *Left panels:* Conditional probability distribution $P(\mu_{50,r}|M_*)$ when using stellar mass estimations based on the fits to $g - i$ colours and absolute magnitudes $M_i$ from Zibetti et al. (2009). Empty circles show the resulting distribution from observations of galaxies with Sersic index $n_s \leq 2$ only. Solid lines show the fit to observations when assuming a lognormal model distribution as described in the text. *Right panels:* SB-to-galaxy stellar mass relation for three different stellar mass estimators used in this paper. Solid circles show the mean values of $\mu_{50}$ that are affected by SB incompleteness while empty circles show that are complete according Blanton et al. (2005b).

$M_* = 10^9 M_\odot$ (where the missing number of low SB galaxies is negligible) as

$$\langle \mu_{50,r} \rangle = \alpha \log M_* + \beta. \tag{C3}$$

For simplicity, we assume that the dispersion around this relation, $\sigma_\mu$, is independent of mass and has the same value for all the mass estimators; we assign a value of 0.75 dex, which is close to most of the values determined by fitting Equation (10) to the data from our galaxy sample for the three methods of assigning stellar masses. Note that the values of $\alpha$ and $\beta$ depend on each stellar mass estimator implying that SB corrections are susceptive to systematics due to stellar masses estimators. The next step is to assume that the distribution of real galaxies, $N_{\text{real}}$, can be generated from the probability distribution $P(\mu_{50}|M_*)$ by simply extrapolating Equations (11) and (10) up to $M_* \sim 10^7$ $M_\odot$. Using the definition of $w_{\mu,j}$ (the SB completeness factor), the observed distribution of galaxies, $N_{\text{obs}}$, is thus generated from the probability distribution $P_{\text{obs}}(\mu_{50,r}|M_*) = (1/w_\mu) \times P(\mu_{50,r}|M_*)$. The factor of missing galaxies below the SB $\mu_{50} = 24$ mag arcsec$^{-2}$ at a given stellar mass is then

$$f_{\text{loss}}(M_*) = \int \left(1/w_\mu\right) P(\mu_{50,r}|M_*) d\mu_{50,r} \Big/ \int P(\mu_{50,r}|M_*) d\mu_{50,r}. \tag{C4}$$

Thus, we weight every galaxy in the sample with:

$$w_{\text{SB},j} = w_{\mu,j} \times w_{\text{loss}}, \tag{C5}$$

where

$$w_{\text{loss}} = \begin{cases} 1/f_{\text{loss}} & \text{for} \quad n_s \leq 2 \\ 1 & \text{for} \quad \text{else.} \end{cases} \tag{C6}$$

We are now in position to estimate the GSMF corrected by SB incompleteness.

### C.2. The dependence of a stellar mass limit sample with redshift

In order to calculate the GSMF, we start by determining how the apparent magnitude limit of the SDSS transforms into a stellar mass limit. In other words, given that the apparent magnitude limit of the SDSS is $m_{r,\text{lim}} = 17.77$, we compute the equivalent in terms of stellar mass, $M_{*,\text{lim}}$. Following van den Bosch et al.

(2008), we determined the redshift-dependent absolute magnitude limit $M_{r,\text{lim}}^{0.0}$ given the apparent magnitude limit from the SDSS $m_{r,\text{lim}} = 17.77$

$$M_{r,\text{lim}}^{0.0} = m_{r,\text{lim}} - 5 \log D_L(z) - 25 - K_{\langle gr \rangle}(z) + E_r(z), \tag{C7}$$

where $D_L$, $K_{\langle gr \rangle}$, and $E_r$ are functions described in Appendix A. Note that we have emphasised the use of average colours for the $K$−correction because we are interested in the stellar mass limit for all the galaxies. Thus, the above absolute magnitude limit depends both on redshift and colour (van den Bosch et al. 2008). Using the colour-dependent mass-to-light ratio $\Upsilon_r^{Z09}(g - r)$ from Zibetti et al. (2009), we transform $M_{r,\text{lim}}^{0.0}$ into a stellar mass limit

$$M_{*,\text{lim}} = -0.84 + 1.654 \times \langle g - r \rangle^{0.0} - 0.4 \times \left(M_{r,\text{lim}}^{0.0} - 4.64\right). \tag{C8}$$

Finally, we use the mean relationship between colour and stellar mass for blue and red galaxies as well as the fraction of red, $f_R$, and blue galaxies, $f_B$, to compute the average colour–stellar mass relationship as

$$\langle g - r \rangle^{0.0} = f_B(g - r)_B^{0.0} + f_Q(g - r)_Q^{0.0}, \tag{C9}$$

where $(g - r)_B^{0.0}$ and $(g - r)_Q^{0.0}$ are the best-fit models to the mean colour–stellar mass relationships of blue and red galaxies.

We paused here for a moment and described our method to derive $(g - r)_B^{0.0}$ and $(g - r)_Q^{0.0}$. To do so, we use the SDSS DR7 based on the photometric catalogue from Meert et al. (2016). We choose to use this catalogue as contains many more galaxies than the SDSS DR4 and one could derive robust colour distributions. We derived the observed distribution function of galaxy colours as a function of stellar mass, $P_{gr}(g - r|M_*)$, that is the observed distribution of galaxy colours at the range between $(g - r)^{0.0} \pm \Delta(g - r)^{0.0}/2$ and $\log M_* \pm \Delta \log M_*/2$. We divide our space into 20 bins equally spaced for $(g - r)^{0.0}$ between $(g - r)^{0.0} = 0$ and $(g - r)^{0.0} = 1.4$ and into 25 bins equally spaced between $\log M_* = 8.5 - 12$. For galaxy stellar masses, we use our fiducial definition from Appendix A.

We assume that the distribution $P_{gr}(g - r|M_*)$ is bimodal and composed of two Gaussian distributions, this is a good approximation as shown by previous studies (e.g., Baldry et al. 2004, 2006). We associate one of the modes of $P_{gr}(g - r|M_*)$ with the distribution of blue galaxies, denoted by $P_{gr,B}(g - r|M_*)$, while the remaining one with the distribution of red galaxies, denoted

by $P_{gr,R}(g - r|M_*)$. The relation between these distributions is given by:

$$P_{gr,B}(g - r|M_*) = f_B(M_*)P_{gr,B}(g - r|M_*)$$
$$+ f_R(M_*)P_{gr,R}(g - r|M_*). \quad (C10)$$

We assume that $P_{gr,j}(g - r|M_*)$, with $j =$ B or R, is a Gaussian distribution given by:

$$P_{gr,j}(g - r|M_*) = \frac{1}{\sqrt{2\pi\sigma_j^2(M_*)}} \times$$
$$\exp\left[-\frac{((g - r)^{0.0} - (g - r)_j^{0.0}(M_*))^2}{2\sigma_j^2(M_*)}\right], \quad (C11)$$

where $(g - r)_j^{0.0}(M_*)$, with $j =$ B or R, is the mean colour–stellar mass relationship used in Equation (17) and $\sigma_j(M_*)$ is the standard deviation that depends on $M_*$. The functional forms for $(g - r)_j^{0.0}(M_*)$ and $\sigma_j(M_*)$ are given by

$$(g - r)_j^{0.0}(M_*) = \alpha_j + \beta_j \times \log\left(\frac{M_*}{10^{11}M_\odot}\right), \quad (C12)$$

and

$$\sigma_j(M_*) = \lambda_j + \kappa_j \times \log(M_*), \quad (C13)$$

for $j =$ B, R. Finally, for the fraction of red galaxies, we assume that

$$f_R(M_*) = \frac{1}{1 + [a + b(M_*/M_C)]^\gamma}. \quad (C14)$$

We performed a $\chi^2$ minimisation procedure to the observed galaxy colour bimodality in order to find the best-fitting parameters to the functional forms described above. Our best-fitting parameters are: $(\alpha_B, \beta_B, \lambda_B, \kappa_B, \alpha_R, \beta_R, \lambda_R, \kappa_R, a, b, \log M_C, \gamma) =$ (0.514, 0.086, 0.240, $-0.015$, 0.720, 0.064, $-0.068$, 0.014, 0.001, 1.390, 10.586, $-1.001$).

The upper panel of Figure C3 shows the dependence of $M_{*,\text{lim}}$ with redshift. The region above $M_{*,\text{lim}}$ is the area above which the NYU-VAGC galaxy sample is a volume-limited sample that is complete in stellar mass. The small grey dots show individual galaxies from the NYU-VAGC sample in the case that stellar masses were determined using the geometric mean of all our stellar mass estimators.

### C.3. Volume and large-scale structure corrections

In a volume-limited sample that is complete in stellar mass, we can derive the GSMF as the number of observed galaxies, $N_{\text{gals}}$, per unit of comoving volume $V$ with stellar masses between $\log M_* \pm \Delta \log M_*/2$, that is, $\phi_*(M_*)\Delta \log M_* = N_{\text{gals}}/V$. Once we determined the dependence of the stellar mass limit with redshift, $M_{*,\text{lim}}$, from the SDSS NYU VAGC sample, we can use the above idea by defining various volume-limited subsamples that are complete in stellar mass. These subsamples were defined by dividing the galaxy redshift range covered by the NYU VAGC, $0.0033 < z < 0.005$, into 20 bins. Therefore, the GSMF for the $j$th volume-limited subsample at the redshift range $z \pm \Delta z/2$ that is complete in stellar mass can be estimated for the mass bin $\log M_* \pm \Delta \log M_*/2$ as

$$\phi_j(M_*, z)\Delta \log M_* = \frac{N_{\text{gals},j}(M_*, z)}{V(z - \Delta z/2) - V(z + \Delta z/2)}. \quad (C15)$$
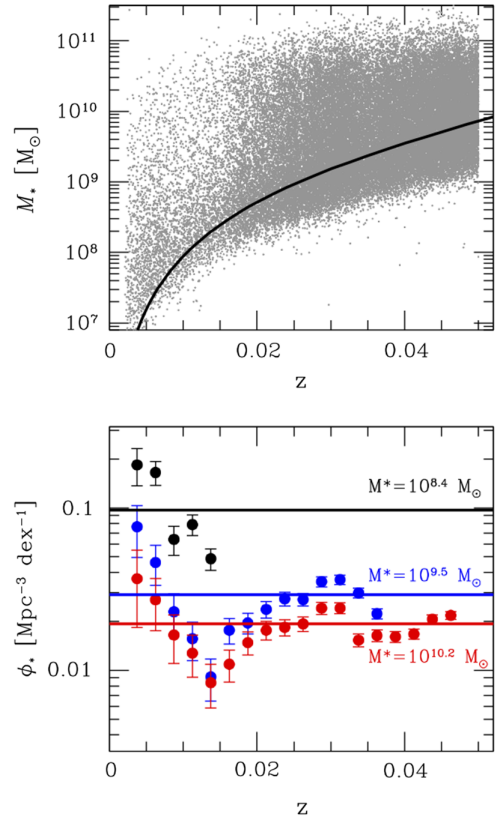


**Figure C2.** *Upper Panel:* Distribution of galaxies in the $M_*$ and redshift plane for the lowz-SDSS galaxy sample, grey dots. The solid lines shows the dependence of the stellar mass completeness limit as a function of redshift for our galaxy sample. *Bottom Panel:* The dependence of the GSMF with redshift for three different stellar masses. Note that the increase and decreases in the amplitudes is due the large-scale structure fluctuations. The solid lines show our corrections due to large-scale structures as described in the text.

We tested the above methodology with realistic mock galaxy catalogues. To do so, we use the $N$-body Bolshoi–Planck simulation (Klypin et al. 2016) and halo catalogues described in Rodríguez-Puebla et al. (2016). We use the semi-emperical modelling from (Rodríguez-Puebla et al. 2017) in order to assign galaxies to dark matter halos/subhalos. The galaxies in the catalogue were projected into the redshift space through a lightcone. We use the dependence of the stellar mass limit with redshift described in Appendix C.2 and include galaxies only within the same redshift range as the NYU VAGC in order to reproduce the observed distribution of galaxies in the $M_*$ and redshift plane for the low-z-SDSS sample. Our results show that the above methodology recovered the original GSMF with differences not larger than ∼5%. In addition, we have calculated the GSMF using the stepwise maximum likelihood method (Efstathiou, Ellis, & Peterson 1988) and found very similar similar results (not shown) as those reported based on our methodology.

Finally, we calculate the GSMF corrected by surface brightness incompleteness by

$$\phi_{\text{SB},j}(M_*, z)\Delta \log M_* = \frac{N_{\text{SB,gals},j}(M_*, z)}{V(z - \Delta z/2) - V(z + \Delta z/2)}, \quad (C16)$$
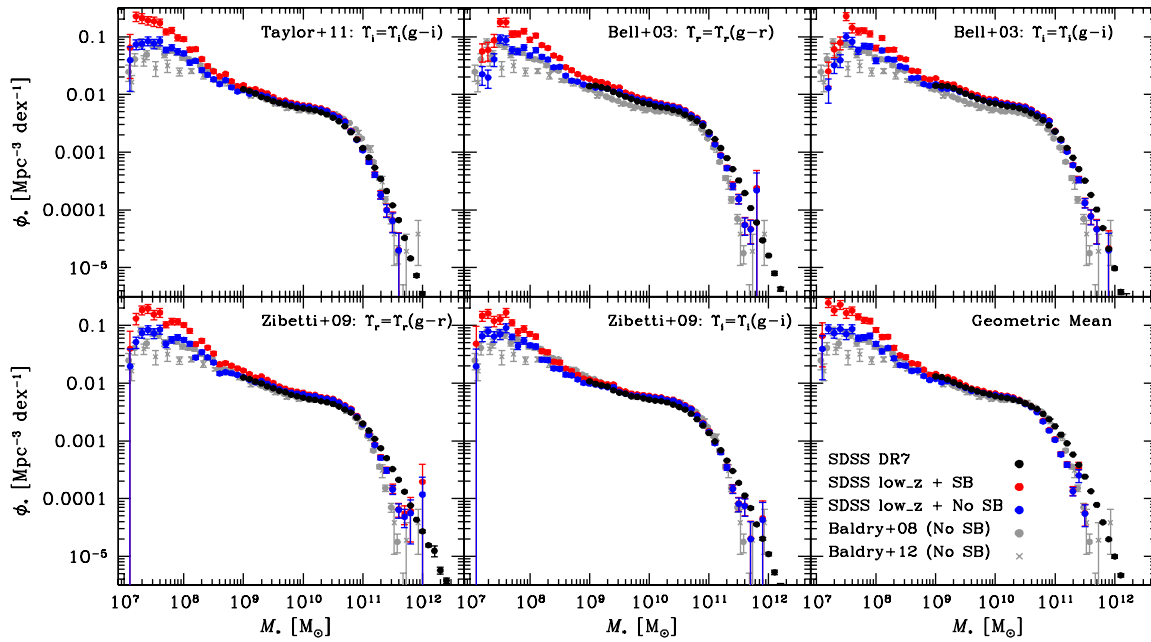
**Figure C3.** Corrected GSMF for surface brightness incompleteness, red filled circles with error bars. Blue filled circles with error bars show the uncorrected GSMF. Grey skeletal symbols and filled circles show the GSMF from Baldry et al. (2012) and Baldry et al. (2008), respectively. The black filled circles show the GSMF from the SDSS DR7 derived from Appendix A. The different panels show the various stellar mass estimators used for this paper. Note that the at $M_* \sim 10^9$ M$_\odot$ there is a smooth transition between the low-*z* NYU-VAGC sample and the SDSS DR7 GSMFs.

where

$$N_{\mathrm{SB,gals},j} = \sum_{i=1}^{N_{\mathrm{gals},j}} w_{\mathrm{SB},i}, \qquad (\mathrm{C17})$$

and $w_{\mathrm{SB},i}$ is our SB incompleteness correction given by Equation (13).

The bottom panel of Figure C3 shows $\phi_{\mathrm{SB},j}(M_*, z_j)$ for three different stellar masses, $M_* = 10^{8.4}$M$_\odot$, $10^{9.5}$M$_\odot$, and $10^{10.2}$M$_\odot$. Fluctuations in the amplitude of the GSMF shows that the distributions of galaxies is not uniform across the redshift distribution because of environmental effects arising from large-scale structures. In order to minimise the above effect, we compute the weighted mean of the GSMF. In other words, we derive the total GSMF as

$$\langle \phi_{\mathrm{SB}}(M_*) \rangle = \sum_{j=1}^{N=20} \phi_{\mathrm{SB},j}(M_*, z_j) \times w_j, \qquad (\mathrm{C18})$$

where $w_j = N_{\mathrm{gals},j} / \sum_j N_{\mathrm{gals},j}$ the fraction of galaxies at the *j*th volume-limited subsample centred at the redshift bin $z \pm \Delta z/2$ for the mass bin $\log M_* \pm \Delta \log M_*/2$. The solid line in Figure C3 shows the resulting value of $\langle \phi_{\mathrm{SB}} \rangle$ for the masses discussed above.

Figure C4 compares the resulting GSMFs when SB corrections are applied $\phi_{\mathrm{SB},*}$ (red filled circles) and when we ignore SB corrections $\phi_*$ (blue filled circles) for each of the six stellar mass definitions used here. As expected, the SB correction increases the number density of low-mass galaxies. For higher masses than $\sim 3 \times 10^9$ M$_\odot$, this correction is negligible. For comparison, we reproduce with grey filled circles the GSMF reported in Baldry et al. (2008), who used also the low-*z* NYU-VAGC sample but for the DR4 as well as the Baldry et al. (2012) from the GAMA survey with the skeletal symbols. In none of them, SB corrections were applied. Finally, in all the panels of Figure C4, we reproduce the

GSMF from the main SDSS DR7 derived in Appendix A. Observe how the GSMFs constructed from the low-z NYU-VAGC sample and the ones constructed from the main SDSS DR7 samples match extremely well at $M_* \sim 10^9$ M$_\odot$, but the latter overcomes the former at high stellar masses due to the larger volume covered by the SDSS DR7.

Finally, we briefly describe our final GSMF. For galaxies below $M_* = 10^9$M$_\odot$, we use the GSMF derived from the low-z NYU-VAGC sample, while for galaxies above $M_* = 10^9$M$_\odot$, we use the GSMF from the SDSS DR7 based on the photometric catalogue from Meert et al. (2015). We apply a simply correction in our GSMF for passing from the Willick et al. (1997) distance flow model to the Tonry et al. (2000) one. Figure C1 from Baldry et al. (2012) shows that after adjusting the Baldry et al. (2008) GSMF to the Tonry et al. (2000) distances, both MFs are in excellent agreement. With that information, we first note that our fiducial (uncorrected) GSMF (bottom right panel from Figure C3) is very similar to the Baldry et al. (2008) GSMF, and thus we assume that the impact of correcting by Tonry et al. (2000) distances is equivalent to rescale it to the Baldry et al. (2012) GSMF. Based on the above, we rescale our SB-corrected GSMF as $\phi_{\mathrm{SB,T00}} = \langle \phi_{\mathrm{SB}} \rangle \times \phi_{\mathrm{B08}}/\phi_{\mathrm{B12}}$. Recall that our fiducial GSMF uses stellar masses from the geometric mean of all stellar masses described by Equation (A6).

Figure C5 shows our final GSMF, $\phi_{\mathrm{SB,T00}}$, as the black filled circles with error bars. The filled grey symbols show the GSMF, $\langle \phi_{\mathrm{SB}} \rangle$, in which the Willick et al. (1997) model flow is utilised. We also compare to Baldry et al. (2012) and Wright et al. (2017) determinations. Note that after distance and SB corrections, our fiducial GSMF is in good agreement with the observed low-mass end slope of the GAMA survey. For comparison, we present our corrected GSMF but when using the Taylor et al. (2011) mass-to-light ratios. Note that in this case our GSMF is consistent with the Baldry et al. (2012) GSMF.

### C.4. The impact of galaxy classification: the criteria for separating the galaxy population into two groups

For our goal of projecting gas scaling correlations (more precisely, the gas CPDFs) into gas MFs separately for early- and late-type galaxies, the derivation of the fraction of early-type galaxies as a function of $M_*$, $f_E(M_*)$, was an important step. As discussed in Section 3.3, based on the morphological classification from Huertas-Company et al. (2011), we found the GSMFs of early- and late-type galaxies that are in good agreement with the results based on the visual classification from Nair & Abraham (2010) and with the classification based on concentration utilised in Bernardi et al. (2010). In contrast, we found that our GSMFs of early- and late-type galaxies are in tension when comparing to those from the GAMA survey with their visual morphological classification, but interestingly enough, they agree with the GAMA GSMFs when we use a $g - r$ colour criterion to separate our galaxies into the two populations. Recall that for the GAMA classification, Sa galaxies are included into their early-type group since their visual classifications combines S0 and Sa galaxies (Kelvin et al. 2014; Moffett et al. 2016a), contrary to our definition; see Section 3.3 for more details. Thus, understanding the impact of using different criteria to separate the galaxy population into two main groups is of great importance in our study. Following, we study the impact of using galaxy colour instead of morphology in order to give a rough idea of what would it be the result of using a very different proxy to galaxy morphology (a one close to the GAMA survey, for instance).

The lower panel of Figure 2 presented the fractions of early-type galaxies as well as of red galaxies as a function of $M_*$. The fraction of red galaxies is clearly larger than the one of early-type galaxies at all masses. Based on the SDSS DR7 sample described in Section 3, we found that the great majority of the galaxies that are classified as early type are actually red; the fraction of early-type galaxies with blue colours has a maximum at $M_* \sim 8 \times 10^{10} M_\odot$ representing only $\sim 5\%$ of the population. In contrast, the fraction of red galaxies classified as late types is larger than $\sim 10\%$ at practically all masses and it peaks at $M_* \sim 2 \times 10^{10} M_\odot$ with a contribution of $\sim 50\%$ (we also observe a second peak at the massive end $M_* \sim 4 \times 10^{11} M_\odot$). Similar results have been reported in previous studies (see e.g., Masters et al. 2010b). Additionally, note that we ignored the effects of reddening due to extinction from the galaxy inclination, which would misclassify galaxies based on their colours (see e.g., Masters et al. 2010a). Therefore, from the physical point of view, the separation of the galaxy population by colour is, perhaps, not as 'clean or reliable as morphology.

Figure C6 presents the resulting MFs when using the fraction of red galaxies, $f_r(M_*)$, as a proxy for early-type galaxies, dashed lines. The solid lines reproduce the results from Figure 8, where our morphology-based fraction, $f_E(M_*)$, was used. Notice that for HI, $H_2$, and cold gas mass, not only the MFs of blue and red galaxies are different to their morphological counterparts but also the total MFs. The above can be understood in the following terms. Using the fraction of red galaxies as a proxy of early-type galaxies results in a large fraction of red galaxies misclassified as late-types as discussed above. However, the above has a lager impact for early-type galaxies with low to intermediate masses than at high masses, while for late-type galaxies, the major impact is from intermediate to high masses. As a consequence, on one hand, the resulting HI and $H_2$ MFs see an increase in early-type galaxies at their low-mass ends. Interestingly enough, the use of $f_r$ instead of $f_E$ would produce HI and $H_2$ MF of early-type galaxies in better agreement
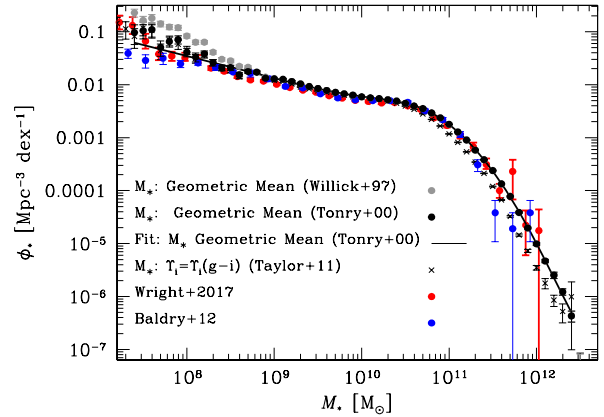


**Figure C4.** Corrected GSMF for SB and for the flow model, filled circles with error bars. This is our fiducial GSMF. The corresponding best-fit model is shown with the solid line, see Section 3.2. Grey filled circles shows when using the flow model from Willick et al. (1997). We also show our results when using the mass-to-light ration from Taylor et al. (2011) and compared to Baldry et al. (2012) who used the same mass estimator. Notice that both mass functions are consistent between each other. For completeness, we compared to Wright et al. (2017).

with the inferences of the ATLAS 3D sample. On the other hand, lowering the fraction of late-type galaxies at intermediate-high masses, which have significantly larger gas fractions than early-type galaxies, affects the projected total HI and $H_2$ MFs, and they would be in tension with direct observations, especially with the HI MF from the ALFALFA and HIPASS surveys.

Finally, we emphasise that the above does not imply that using galaxy colours will lead to incorrect inferences of the gas MFs but that combining two different criteria for dividing the galaxy population will lead to a very different results that, perhaps, will be in tension with the observations. Thus, the success of our determinations is in part that we are using data sets that are consistent between each other in that regards the morphological separation into two galaxy subpopulations.

## D. Deconvolution algorithm

Individual mass estimates are subject to random errors. Thus, every MF that is inferred from observations through indirect estimations of any type of mass (we will denote this as $\phi_{obs}$) is the result of the random errors over the intrinsic mass (it will be denoted by $\phi_{int}$). Formally, we can represent the observed $\phi_{obs}$ as the convolution of $\phi_{int}$:

$$\phi_{obs}(M) = \int \mathcal{G}(\log M - \log x)\phi_{int}(x)d\log x. \quad (D1)$$

We will assume that random errors have a lognormal distribution, denoted by $\mathcal{G}(\log M - \log x)$:

$$\mathcal{G}(\log M - \log x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\log^2\left(\frac{M}{x}\right)\right], \quad (D2)$$

where $\sigma$ are the 1-$\sigma$ statistical fluctuations, in either directions, in the inferred galaxy masses. Note that in Equation (27), the units for $\phi_{obs}$ and $\phi_{int}$ are in Mpc$^{-3}$dex$^{-1}$.

The basic idea of our algorithm is simple. We start by defining the following relation:

$$\phi_{int}^j(M) = \phi_{int}^{j-1}(M) \int \mathcal{G}(\log M - \log x)\frac{\phi_{obs}}{\phi_{conv}^{j-1}}(x)d\log x, \quad (D3)$$
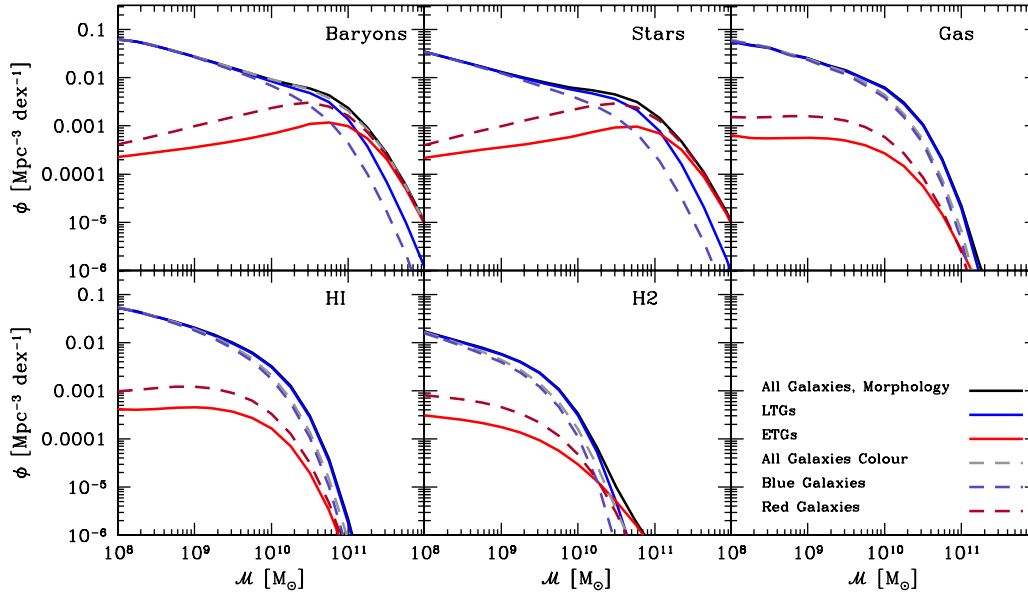
**Figure C5.** Impact on the MFs due to the use of two different criteria for the division of the galaxy population. The solid lines show the original MF from Section 4.4, based on galaxy morphology, while the dashed lines show the results when using galaxy colour. The classification according to galaxy colours results in a overabundance of red galaxies compared to early types, especially at intermediate and low masses.

where

$$\phi_{\text{conv}}^{j-1}(x) = \int \mathcal{G}(\log x - \log y)\phi_{\text{int}}^{j-1}(y)d\log y, \qquad \text{(D4)}$$

with $\phi_{\text{int}}^{j-1}$ denoting the $j$th iterated intrinsic MF. Note that as $\phi_{\text{conv}}^{j-1}$ approaches to $\phi_{\text{obs}}$ the above equation converges to the maximum likelihood solution for $\phi_{\text{int}}^{j-1}$, in other words, we have found the numerical solution to the intrinsic MF, $\phi_{\text{int}}$. The zero-th iteration is defined as convolution of the observed MF with the lognormal distribution $\mathcal{G}$ described above:

$$\phi_{\text{int}}^{0}(M_*) = \int \mathcal{G}(\log M_* - \log x)\phi_{\text{obs}}(x)d\log x. \qquad \text{(D5)}$$

We declare that the $\phi_{\text{int}}^{j}$ has converged when the parameter $\Delta \leq 7\%$ defined as the relative error between the observed MF and the $j$-th iterated intrinsic MF convolved with the random error distribution:

$$\Delta = \frac{100\%}{N} \sum_{i} \left| 1 - \frac{\int \mathcal{G}(\log M_* - \log x)\phi_{\text{int},i}^{j}d\log x}{\phi_{\text{obs},i}} \right|. \qquad \text{(D6)}$$

The summation in the above definition goes over all the tabulated values of individual reports of the observed MF $\phi_{\text{obs}}$. By trial and error, we found that the value of $\Delta = 7\%$ is a compromise between accuracy and efficiency. Typically, $\Delta = 7\%$ was reached in less than 10 iterations.